THE

PHILOSOPHICAL MAGAZINE

EDITOR
PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD
SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.
ALLAN FERGUSON, M.A., D.Sc.
SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.
PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

VOL. XLI.—SEVENTH SERIES

JANUARY—DECEMBER 1950

Reprinted by

Wm. DAWSON & SONS LTD.,

with the permission of

TAYLOR & FRANCIS LTD.

"Meditationis est perscrutari occulta; contemplationis est admirari perspicua.... Admiratio generat quæstionem, quæstio investigationem, investigatio inventionem."—*Hugo de S. Victore.*

—— "Cur spirent venti, cur terra dehiscat,
Cur mare turgescat, pelago cur tantus amaror,
Cur caput obscura Phœbus furrugine condat,
Quid toties diros cogat flagrare cometas,
Quid pariat nubes, veniant cur fulmina cœlo,
Quo micet igne Iris, superos quis conciat orbes
Tam vario motu."

J. B. Pinelli ad Mazonium.

ALERE FLAMMAM.

CONTENTS OF VOL. XLI. (Nos. 312–323)

(SEVENTH SERIES)

NUMBER 312—JANUARY

Page

I. Some Conformal Transformations involving Elliptic Functions. By Sir Charles Darwin, F.R.S., National Physical Laboratory. 1

II. The Propagation of Sound Waves in an Open-ended Channel. By W. Chester 11

III. Betatron-Starting in Electron Synchrotrons. By J. J. Wilkins, Atomic Energy Research Establishment, Harwell. 34

IV. On the Interaction of Colloidal Particles.—III. Some Theorems Concerning the Free Energy of Double Layers. By S. Levine, Ph.D., Birkbeck College Research Laboratory 53

V. Noise in Ionization Chamber Pulse Amplifiers. By R. Wilson, D.Phil., Clarendon Laboratory, Oxford 66

VI. A Radio Echo Method for the Measurement of the Heights of the Reflecting Points of Meteor Trails. By J. A. Clegg and I. A. Davidson, Physical Laboratories, The University, Manchester 77

VII. On the Mass of Charged Particles of the Cosmic Radiation. By C. Franzinetti, D. in F., H. H. Wills Physical Laboratory, University of Bristol. (Plates I. & II.) 86

VIII. Notices of New Books and Periodicals received :—

- Supersonic Flow Around Cones of Large Yaw 107
- Prof. H. Butterfield's The Origins of Modern Science. 107
- Prof. J. R. Partington's An Advanced Treatise on Physical Chemistry.—Vol. I. Fundamental Principles and Properties of Gases 107
- Mr. S. A. Schelkunoff's Applied Mathematics for Engineers and Scientists 108
- Profs. N. F. Mott and H. S. W. Massey's The Theory of Atomic Collisions 108

NUMBER 313—FEBRUARY

IX. Atmospheric Absorption of Heat Radiation by Water Vapour. By T. G. Cowling 109

X. Orthogonal Polynomials in the Least Squares Fitting of Observations. By P. G. Guest, Lecturer in Physics, University of Sydney. 124

XI. Emission of Protons and Alpha-Particles in High Energy Cosmic Ray Stars. By D. H. Perkins, Imperial College of Science and Technology, London, S.W. 7. 138

XII. On the Fine Structure near the FeK-absorption Edges in the Various Oxides of Fe. By D. Coster and S. Kiestra, Natuurkundig Laboratorium der Rijks-Universiteit, Groningen, The Netherlands 144

XIII. Gravitational Fields of Spherical Symmetry and Weyl's Conformal Curvature Tensor. By V. V. Narlikar and K. P. Singh, Benares Hindu University, India 152

XIV. The Fluorescence of Silver Halides at Low Temperatures.—Part I. The Pure Halides. By G. C. Farnell, P. C. Burton and R. Hallama, Research Laboratories, Kodak Limited, Wealdstone, Middlesex. (Plates III.–VI.) 157

MG124
75
3 new

Page
XV. Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part III. Nature of the Shower Particles. By P. H. Fowler, The H. H. Wills Physical Laboratory, University of Bristol.... 169
XVI. The S-Matrix for Meson-Nucleon Interactions. By P. T. Matthews, Clare College, Cambridge 185
XVII. Correspondence :—
Observation of Unimolecular Growth Steps on Crystal Surfaces. By L. J. Griffin, Department of Physics, Royal Holloway College. (Plates VII. & VIII.).... 196
Note on the Structure of a Crystal Surface. By F. C. Frank, H. H. Wills Physical Laboratory, Bristol.... 200
Transfer Rates for Surface Films of HeII. By E. J. Burge and L. C. Jackson, H. H. Wills Physical Laboratory, University of Bristol 205
XVIII. Notices of New Books and Periodicals received :—
S. Tolansky's Introduction to Atomic Physics 208
J. A. Crowther's Ions, Electrons and Ionizing Radiations .. 208
T. Martin's Faraday's Discovery of Electromagnetic Induction.... 208

NUMBER 314—MARCH

XIX. Note on Potentials Derived from Axial Values in Electron Optics. By Feodora Berz, L. ès Sc., Mullard Electronic Research Laboratories .. 209
XX. Properties of Slow Electrons in Polar Materials. By H. Fröhlich, H. Pelzer and S. Zienau, Department of Theoretical Physics, The University, Liverpool 221
XXI. Kinetics of the Phase Transition in Superconductors. By A. B. Pippard, Royal Society Mond Laboratory, Cambridge.... 243
XXII. Programming a Computer for Playing Chess. By Claude E. Shannon, Bell Telephone Laboratories, Inc., Murray Hill, N.J. 256
XXIII. Propagation of Electromagnetic Disturbances along a Thin Wire in a Horizontally Stratified Medium. By B. L. Coleman, B.Sc., British Electrical and Allied Industries Research Association 276
XXIV. Quantal Aspects of Scientific Information. By D. M. MacKay, Wheatstone Laboratory, King's College, London 289
XXV. Notices of New Books and Periodicals received :—
R. N. Haward's The Strength of Plastics and Glass 312
H. R. Kruyt's Colloid Science.... 312

NUMBER 315—APRIL

XXVI. Theory of Elasticity in General Coordinates. By A. E. Green and W. Zerna, Department of Mathematics, King's College, Newcastle-upon-Tyne 313
XXVII. The Emission of Short-range Alpha Particles from Light Elements under Proton Bombardment.—II. Further Observations on the Reaction $^{15}B(px)^{7}Be$. By W. E. Burcham and Joan M. Freeman 337
XXVIII. Relaxation Methods applied to Problems of Plastic Flow.—I. Notched Bar under Tension. By J. A. Jacobs, Royal Holloway College, University of London 349

Page

XXIX. On Tolman's Equation describing the Thermal Equilibrium in a Gravitating Sphere of Fluid. By H. A. Buchdahl, Physics Department, University of Tasmania 362

XXX. Scattering of Pseudoscalar Charged Mesons by Nucleons.—II. By E. Corinaldesi and G. Field, Dublin Institute for Advanced Studies. 364

XXXI. Influence of the Overlap Integral in Molecular-orbital Calculations for Alternant Hydrocarbons. By J. de Heer, Wheatstone Physics Department, King's College, London 370

XXXII. The Estimation of Velocities and Local Mach Numbers in the Flow of a Compressible Fluid through Divergent Nozzles, under Adiabatic and Isothermal Conditions. By R. K. Tempest 382

XXXIII. Noise in Current-carrying Ohmic Conductors. By Bernard Meltzer 393

XXXIV. A 4-dimensional Generalization of Wilson's Hypothesis. By A. Papapetrou, I.C.I. Research Fellow, University of Manchester.... 399

XXXV. Correspondence:—

- Further Evidence for the Existence of τ-Mesons. By J. B. Harding, Imperial College, London........ 405
- Intrinsic Block Structure and the Strength of Metals. By M. S. Paterson, Ph.D., Aeronautical Research Laboratories, Melbourne........ 410

XXXVI. Notices of New Books and Periodicals received:—

- L. M. Milne-Thomson's Theoretical Hydrodynamics 411
- H. T. Jessop and F. C. Harris's Photoelasticity 411
- Norbert Wiener's The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications 411
- L. R. G. Treloar's The Physics of Rubber Elasticity...... 412
- Markus Reiner's Twelve Lectures on Theoretical Rheology. 412

NUMBER 316—MAY

XXXVII. Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part IV. The Distribution in Energy, and the Secondary Interactions of the Particles Emitted from Stars. By U. Camerini, P. H. Fowler, W. O. Lock, and H. Muirhead, The H. H. Wills Physical Laboratory, University of Bristol. (Plates IX.–XI.) 413

XXXVIII. Energy Spectrum of Fragments from the Spontaneous Fission of Natural Uranium. By W. J. Whitehouse and W. Galbraith 429

XXXIX. Scattering of 90 MeV. Neutrons by Alpha-Particles. By J. Heidmann, Ecole Polytechnique, Paris........ 444

XL. Relaxation Methods applied to Problems of Plastic Flow.—II. By J. A. Jacobs, M.A., Ph.D., A.F.R.Ae.S., Royal Holloway College, University of London 458

XLI. Some Extensions in the use of Resistance Thermometry in the Study of Gaseous Explosions. By A. Smeeton Leah, Ph.D., A.M.I.Mech.E., C. Rounthwaite, B.Sc., and D. Bradley, Ph.D. (Plate XII.) 468

XLII. Atomic Oxygen in Carbon-Monoxide Explosions. By A. Smeeton Leah, Ph.D., A.M.I.Mech.E., C. Rounthwaite, B.Sc., and D. Bradley, Ph.D........ 478

Page

XLIII. The Cut-Off in the Energy Spectrum of Primary Cosmic Radiation. By P. J. D. Gething, Department of Mathematics, Imperial College of Science, London 495

XLIV. Correspondence :—

On the Binary and Ternary Photofission of Thorium-232. By E. W. Titterton and T. A. Brinkley 500

A Temporary Increase in the Neutron Component of Cosmic Rays. By N. Adams, The Physical Laboratories, The University, Manchester, 13 503

A Temporary Increase in the Neutron Component of Cosmic Rays. By N. Adams and H. J. J. Braddick 505

XLV. Notices of New Books and Periodicals received :—

M. Born and H. S. Green's A General Kinetic Theory of Liquids 506

Claude E. Shannon and Warren Weaver's The Mathematical Theory of Communication 508

Gerard P. Kuiper's The Atmospheres of the Earth and Planets. 508

NUMBER 317—JUNE

XLVI. The Refractive Index and Classical Radiative Processes in an Ionized Gas. By K. C. Westfold, The Queen's College, Oxford 509

XLVII. Investigation of Soft Radiations.—VII. Energy Expenditure per Ion-pair for Slow Electrons in Various Gases. By S. C. Curran, A. L. Cockroft, and G. M. Insch, Department of Natural Philosophy, University of Glasgow 517

XLVIII. Magnetic Spectrometer of Large Solid Angle. By E. R. Rae, Department of Natural Philosophy, University of Glasgow. (Plate XIII.) 525

XLIX. The Influence of Exchange and Correlation Forces on the Specific Heat of Free Electrons in Metals. By E. P. Wohlfarth, Ph.D., Department of Mathematics, Imperial College, London 534

L. Note on the Inversion of the Laplace Transform. By B. Gross 543

LI. The Fluorescence of Silver Halides at Low Temperatures.—Part II. Mixed Crystals of Silver Halides. By G. C. Farnell, Ph.D., P. C. Burton, B.Sc., and R. Hallama. (Plates XIV.–XVII.) 545

LII. A Determination of the Energy Distribution of Electrons in the Positive Columns of Gas Discharges. By C. Haigh and T. Smith, Physics Department, Leeds University 557

LIII. The Cohesive Energy of Metallic Magnesium. By S. Raimes, B.Sc., A.R.C.S., D.I.C., Department of Mathematics, Imperial College, London 568

LIV. Nuclear Reactions Produced by Slow Negative π-Mesons. By M. G. K. Menon, H. Muirhead, and O. Rochat, The H. H. Wills Physical Laboratory, University of Bristol. (Plates XVIII. & XIX.) 583

LV. Notices of New Books and Periodicals received :—

P. W. Bridgman's The Nature of Physical Theory 619

John F. Rider's Electronic Engineering Master Index, 1947–48 619

W. E. Milne's Numerical Calculus 619

H. T. Jessop and F. C. Harris's Photoelasticity 620

The Computation Laboratory's Tables of Inverse Hyperbolic Functions 620

Cornelius Lanczos's The Variational Principles of Mechanics .. 620

NUMBER 318—JULY

Page

LVI. The Thermal Conductivity of Potassium Chrome Alum at Temperatures below One Degree Absolute. By C. G. B. Garrett, Royal Society Mond Laboratory, Cambridge.......... 621

LVII. The Crystallography and Thermodynamics of Order–Disorder Transitions in Certain Mixed Ketones. By Vera Daniel, British Electrical Research Association. (Plates XX. & XXI.) 631

LVIII. On a Fourth-Order Meson-Equation. By W. Thirring, Dublin Institute for Advanced Studies.......... 653

LIX. The Effect of Electron Concentration on the Lattice Spacings in Magnesium Solid Solutions. By H. Jones, Mathematics Department, Imperial College, London 663

LX. On the Correlation of the Directional Properties of Rolled Sheet in Tension and Cupping Tests. By L. Bourne, B.Met., and R. Hill, M.A., Ph.D., Metal Flow Research Laboratory, Sheffield 671

LXI. The Fluctuation and Fading of Radio Echoes from Meteor Trails. By J. S. Greenhow, Univeristy of Manchester. (Plates XXII. & XXIII.) 682

LXII. The Influence of High Altitude Winds on Meteor Trail Ionization. By C. D. Ellyett, Ph.D., University of Manchester.......... 694

LXIII. Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part V. The Neutral Mesons. By A. G. Carlson, J. E. Hooper and D. T. King, The H. H. Wills Physical Laboratory, University of Bristol. (Plate XXIV.) 701

LXIV. Correspondence :—

- Gap Measurement as a Method of Analysing Cosmic Ray Stars in Emulsions. By P. E. Hodgson.......... 725
- A Note on the Use of Resistance Thermometers for Measurement of Rapidly Changing Temperatures. By C. B. Daish, D. H. Fender and A. J. Woodall, Military College of Science, Physics Branch 729

LXV. Notices of New Books and Periodicals received :—

- Relations entre les phénomenes solaires et géophysiques. Méthodes de calcul dans des problémes de mécanique. Cinétique et mécanisme des réactions d'inflammation et de combustion en phase gazeuse 730
- N. V. Sidgwick's The Chemical Elements and their Compounds.......... 731
- J. J. Stoker's Non-Linear Vibrations in Mechanical and Electrical Systems.......... 731
- Yvtete Cauchois's Les spectres de rayons X et la structure électronique de la matière 731
- E. G. Richardson's Dynamics of real Fluids 732
- M. Fishenden and O. A. Saunders's An Introduction to Heat Transfer 732

NUMBER 319—AUGUST

LXVI. On the Inhomogeneous Deformation of a Plastic Lamina in a Compression Test. By R. Hill, M.A., Ph.D., Metal Flow Research Laboratory, Sheffield 733

LXVII. A Theoretical Investigation of the Effect of Specimen Size in the Measurement of Hardness. By R. Hill, M.A., Ph.D., Metal Flow Research Laboratory, Sheffield.......... 745

Page

LXVIII. On the Back Current in Blocking-Layer Rectifiers. By J. H. Gisolf 754

LXIX. The Electrical Conductivity of Bismuth Fibres.—II. Anomalies in the Magneto-Resistance. By B. Donovan and G. K. T. Conn, Department of Physics, The University of Sheffield 770

LXX. The Experimental Determination of the Geomagnetic Radial Variation. By S. K. Runcorn, A. C. Benson and A. F. Moore, The Department of Geodesy and Geophysics, Cambridge, and D. H. Griffiths, University, Manchester 783

LXXI. Energies and Widths of Domain Boundaries in Ferromagnetics. By B. A. Lilley, Physics Department, University of Leeds 792

LXXII. The Statistical Analysis of Electrical Noise. By D. K. C. MacDonald, Clarendon Laboratory, Oxford 814

LXXIII. The Origin of Single Cosmic Ray Protons. By S. Lattimore, Imperial College of Science and Technology, London 819

LXXIV. The Control of a Wilson Cloud Chamber by means of an Internal Counter. By A. L. Hodson, A. Loria and N. V. Ryder, The Physical Laboratories, The University, Manchester. (Plates XXV.–XXVIII.) 826

LXXV. The Nature of the Neutral Particle Emitted in the Decay of the π-Meson. By C. O'Ceallaigh, The H.H. Wills Physical Laboratory, University of Bristol. (Plate XXIX.) 838

LXXVI. Correspondence:—

- An Estimate for the Position of the Lowest Dipole-level of a Nucleus. By B. F. Touschek, Department of Natural Philosophy, The University, Glasgow 849
- Examples on the Calculation of Energy States by the Uncertainty Relations. By Hsin P. Soh, National Chekiang University 851

LXXVII. Notices of New Books and Periodicals received:—

- S. Chandrasekhar's Radiative Transfer 855
- G. F. Lothian's Absorption Spectrophotometry 855
- L. Jánossy's Cosmic Rays 856
- Max Von Laue's History of Physics 856
- G. Dyson's Manual of Organic Chemistry for Advanced Students 856

NUMBER 320—SEPTEMBER

LXXVIII. Calibration of Proportional Counters by the Excitation of Fluorescence Radiation with Radioactive Sources. By G. M. Insch, Department of Natural Philosophy, Glasgow University 857

LXXIX. Transit-time Phenomena in Electron Streams.—II. By D. K. C. MacDonald, Clarendon Laboratory, Parks Road, Oxford 863

LXXX. Proportional Counters in a Magnetic Field: an Investigation of the Isomerism of Br^{80} and a Measurement of the Fluorescence Yield of Krypton. By D. West and Miss P. Rothwell, Atomic Energy Research Establishment, Harwell 873

LXXXI. The Eddy Viscosity in Turbulent Shear Flow. By A. A. Townsend, Emmanuel College, Cambridge 890

LXXXII. The Effect of Eddy Viscosity on Ocean Waves. By K. F. Bowden, Oceanography Department, University of Liverpool 907

Page

LXXXIII. Search for the Negative Proton in Fission. By J. M. C. Scott and E. W. Titterton, Atomic Energy Research Establishment, Harwell 918

LXXXIV. The Emission of Short Range Alpha Particles from Light Elements under Proton Bombardment.—III. The Reactions ^{6}Li($p\alpha$)^{3}He ^{7}Li($p\gamma$)^{8}Be. By W. E. Burcham and Joan M. Freeman, Cavendish Laboratory, Cambridge.......... 921

LXXXV. Short Range Alpha Particles from the Break-up of ^{8}Li. By A. S. Baxter, W. E. Burcham and E. B. Paul, Cavendish Laboratory, Cambridge.......... 937

LXXXVI. The Energy Release in the Reaction ^{6}Li(dp)^{8}Li. By E. B. Paul, Cavendish Laboratory, Cambridge 942

LXXXVII. On the Quantum Mechanics of Fluids. By P. J. Price, Royal Society Mond Laboratory, Cambridge 948

LXXXVIII. Correspondence :—

Latitude Effect of Cosmic Ray Stars. By S. Lattimore, Imperial College of Science and Technology, London 961

Note on the Heavy Nuclei of the Cosmic Radiation. By A. D. Dainton and D. W. Kent, H. H. Wills Physical Laboratory, University of Bristol. (Plates XXX. and XXXI.) 963

Loss of Energy of Fast Protons in Matter with Particular Reference to their Range in Carbon. By A. E. Taylor, Atomic Energy Research Establishment, Harwell, Berks. 966

NUMBER 321—OCTOBER

LXXXIX. The Statistics of Correlated Events.—I. By C. Domb, Clarendon Laboratory, Oxford 969

XC. The Di-Iso-Electronic Row and the Effective Charge in the L Isorows. By Anthony Felix Richter, Caroline University, Prague . 983

XCI. The Laminar Boundary Layer on Solids of Revolution. By L. G. Whitehead, Ph.D., and G. S. Canetti, Ph.D., Aeronautical Department, Queen Mary College, University of London 988

XCII. A Theory of Radar Information. By P. M. Woodward, B.A., and I. L. Davies, M.A., T.R.E., Ministry of Supply 1001

XCIII. Some Optical Properties of Evaporated Layers of Silver, Copper and Tin. By D. G. Avery, Ph.D., Royal Holloway College, London University. (Plates XXXII. & XXXIII.) 1018

XCIV. The Magnetic Deflection of Fast Charged Particles in the Photographic Emulsion. By C. C. Dilworth, S. J. Goldsack, Y. Goldschmidt-Clermont and F. Levy. (Plate XXXIV.).......... 1032

XCV. The Momentum and Sign of Fast Cosmic Ray Particles. By J. E. Moyal, Department of Mathematics, The University, Manchester.. 1058

XCVI. Correspondence :—

The Velocity of Second Sound Below 1° K. By K. R. Atkins and D. V. Osborne, Royal Society Mond Lab., Cambridge.......... 1078

The Velocity of Flow of Helium II in the Creeping Film. By L. C. Jackson and D. G. Henshaw, H.H. Wills Physical Laboratory, University of Bristol.......... 1081

NUMBER 322—NOVEMBER

Page

XCVII. Deformation Textures in Face-Centred Cubic Metals. By E. A. CALNAN, B.Sc., and C. J. B. CLEWS, Ph.D., National Physical Laboratory, Metallurgy Division 1085

XCVIII. The Energy Distribution behind Decaying Shocks.—I. Plane Waves. By M. J. LIGHTHILL 1101

XCIX. Energy Losses of Ionizing Particles at Relativistic Velocities in the Photographic Plate. By H. MESSEL and D. M. RITSON, Dublin Institute for Advanced Studies 1129

C. A Theory of the Plastic Bulging of a Metal Diaphragm by Lateral Pressure. By R. HILL, M.A., Ph.D., Metal Flow Research Laboratory, Sheffield 1133

CI. A Method for the Mapping of Vector Potential Distributions in Axially Symmetrical Systems. By G. LIEBMANN, Associated Electrical Industries Research Laboratory, Aldermaston, Berks 1143

CII. Double Stars due to Cosmic Rays. By T. T. LI, Imperial College, London 1152

CIII. Communication Theory and Physics. By D. GABOR, Imperial College, London 1161

CIV. Density Variations in Aluminium. By J. L. SNOEK, Philips Research Laboratories, Eindhoven 1188

CV. Correspondence :—

- Radiative Corrections in the Non-Relativistic Limit. By W. THIRRING, D.Phil., Dublin Institute for Advanced Studies. 1193
- Multiple Processes in the Production of Pairs of Electrons by γ-Radiation. By J. E. HOOPER and D. T. KING, H. H. Wills Physical Laboratory. (Plate XXXV.) 1194

CVI. Notices of New Books and Periodicals received :—

- H. DINGLE's Practical Applications of Spectrum Analysis .. 1196
- D. H. BANGHAM's Progress in Coal Science 1197
- MARTIN JOHNSON's Astronomy of Stellar Energy and Decay .. 1197
- D. H. BANGHAM's Progress in Coal Science 1197

NUMBER 323—DECEMBER

CVII. The Radioactivity of F^{20}.—I. By J. V. JELLEY, Cavendish Laboratory, Cambridge 1199

CVIII. The Radioactivity of F^{20}.—II. By R. M. LITTAUER, Cavendish Laboratory, Cambridge 1214

CIX. The Emission of Short Range Alpha Particles from Light Elements under Proton Bombardment.—IV. The Reaction $^{19}F(p\alpha)^{16}O$. By JOAN M. FREEMAN, Cavendish Laboratory, Cambridge 1225

CX. An Interferometric Study of some Optical Properties of Evaporated Silver Films. By R. C. FAUST, Ph.D., Royal Holloway College. (Plates XXXVI. & XXXVII.) 1238

CXI. The "Velocity" of Brownian Movement. By R. FÜRTH, Birkbeck College, University of London 1255

CXII. The Use of Atmospherics to Study the Propagation of Very Long Radio Waves. By F. F. GARDNER, Cavendish Laboratory, Cambridge 1259

CXIII. The Force Acting on a Body Moving Uniformly through a Gas containing Sound Waves. By F. R. N. NABARRO, Department of Metallurgy, The University of Birmingham 1270

Page
CXIV. A Note on the Decay of Current in Germanium Diodes. By J. R. Tillman and H. Yemm, The Post Office Engineering Research Station, London 1281
CXV. Correspondence :—
Superconductivity of Niobium. By L. C. Jackson and H. Preston-Thomas, H. H. Wills Physical Laboratory, University of Bristol........ 1284
Lattice Sums for Ionic Crystals. By F. C. Frank, The H. H. Wills Physical Laboratory, University of Bristol.. 1287
Some Extensions in Use of Resistance Thermometry in the Study of Gaseous Explosions. By A. S. Leah, C. Rounthwaite and D. Bradley, University of Leeds 1289
CXVI. Notices of New Books and Periodicals received :—
J. K. Syrkin and M. E. Dyatkina's Structure of Molecules and the Chemical Bond........ 1292
Index to Volume 1293

PLATES

I. & II. Illustrative of C. Franzinetti's Paper on the Mass of Charged Particles of the Cosmic Radiation.
III.–VI. Illustrative of G. C. Farnell, P. C. Burton and R. Hallama's Paper on the Fluorescence of Silver Halides at Low Temperatures.—Part I. The Pure Halides.
VII. & VIII. Illustrative of L. J. Griffin's Correspondence on the Observation of Unimolecular Growth Steps on Crystal Surfaces.
IX.–XI. Illustrative of U. Camerini, P. H. Fowler, W. O. Lock and H. Muirhead's Paper on Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part IV. The Distribution in Energy, and the Secondary Interactions of the Particles Emitted from Stars.
XII. Illustrative of A. Smeeton Leah, C. Rounthwaite and D. Bradley's Paper on Some Extensions in the use of Resistance Thermometry in the Study of Gaseous Explosions.
XIII. Illustrative of E. R. Rae's Paper on the Magnetic Spectrometer of Large Solid Angle.
XIV.–XVII. Illustrative of G. C. Farnell, P. C. Burton and R. Hallama's Paper on the Fluorescence of Silver Halides at Low Temperatures.—Part II. Mixed Crystals of Silver Halides.
XVIII. & XIX. Illustrative of M. G. K. Menon, H. Muirhead and O. Rochat's Paper on Nuclear Reactions Produced by Slow Negative π-Mesons.
XX. & XXI. Illustrative of Vera Daniel's Paper on the Crystallography and Thermodynamics of Order-Disorder Transitions in Certain Mixed Ketones.

XXII. & XXIII. Illustrative of J. S. Greenhow's Paper on the Fluctuation and Fading of Radio Echoes from Meteor Trails.

XXIV. Illustrative of A. G. Carlson, J. E. Hooper and D. T. King's Paper on Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part V. The Neutral Mesons.

XXV.–XXVIII. Illustrative of A. L. Hodson, A. Loria and N. V. Ryder's Paper on the Control of a Wilson Cloud Chamber by means of an Internal Counter.

XXIX. Illustrative of C. O'Ceallaigh's Paper on the Nature of the Neutral Particle Emitted in the Decay of the π-Meson.

XXX. & XXXI. Illustrative of A. D. Dainton and D. W. Kent's Correspondence on the Heavy Nuclei of the Cosmic Radiation.

XXXII. & XXXIII. Illustrative of D. G. Avery's Paper on Some Optical Properties of Evaporated Layers of Silver, Copper and Tin.

XXXIV. Illustrative of C. C. Dilworth, S. J. Goldsack, Y. Goldschmidt-Clermont and F. Levy's Paper on the Magnetic Deflection of Fast Charged Particles in the Photographic Emulsion.

XXXV. Illustrative of J. E. Hooper and D. T. King's Correspondence on the Multiple Processes in the Production of Pairs of Electrons by γ-Radiation.

XXXVI. & XXXVII. Illustrative of R. C. Faust's Paper on an Interferometric Study of Some Optical Properties of Evaporated Silver Films.

THE

PHILOSOPHICAL MAGAZINE

A JOURNAL OF THEORETICAL EXPERIMENTAL
AND APPLIED PHYSICS

First published in 1798

[SEVENTH SERIES—VOL. 41]

I. *Some Conformal Transformations involving Elliptic Functions.*

By Sir CHARLES DARWIN, F.R.S.,
National Physical Laboratory*.

[Received September 19, 1949.]

ABSTRACT.

Conformal transformations are given which are suitable for the discussion of the space outside a *finned rectangle*. These include figures such as rectangle, cross, three-sided box, form like the letter H, "windmill", and the transformation could be applied to any rectangle with four fins of any lengths at its corners, each of them in either of two directions. The transformations convert the outer space into the inside of a rectangle, which can then be transformed back into the outside of a circle, with the same point at infinity as the original space.

By a generalization of a special case of this transformation another is given, whereby an irregular star, composed of any number of rays of any lengths and at any angles branching out from a centre, can be transformed into a circle. This includes, as a simple example, a "dihedral" of two equal arms at any angle.

1. THERE are many uses to which conformal transformations are put in physical problems, and the discovery of the appropriate transformations is not always easy. The only general systematic method of discovering them is the Schwartz Christoffel method, and this is limited to cases

---

* Communicated by the Author.

where only half the plane is of interest, and to cases where this half-plane can by symmetry be extended to describe the whole plane. As soon as less symmetrical shapes are considered, the discovery of the appropriate transformations becomes guess-work, and it may therefore be useful to put on record a set of such transformations representing a rather wide range of forms.

The present note started with some work of V. M. Falkner* in connection with finned aeroplanes, for which I provided him with certain transformations involving elliptic functions. For this work the Schwartzian method sufficed, but I have had the curiosity to extend it further into a range of shapes for which that method cannot be used. It is hardly necessary to say that, though it was aerodynamics that gave rise to these problems, they are not to be interpreted as giving directly the air flow over an aeroplane of the shape described.

Elliptic functions will rationalize four square roots, and therefore any form involving four right-angled turns can be represented by their means. There is, however, one important further limitation on the shapes here considered, in that I shall only consider the transformations mapping the area *outside* the form and extending to infinity in all directions.

Fig. 1.

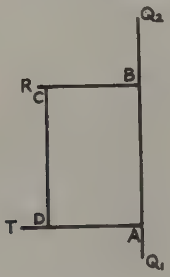
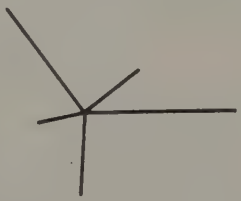


Fig. 2.



The present work gives the transformations whereby the area outside any *finned rectangle* like that in fig. 1 can be transformed into the inside of a rectangle. This rectangle can then in turn be transformed into the area outside a circle in such a manner that the point at infinity coincides with the original point at infinity. In the course of this work the consideration of one special case, the cross, suggested another generalization by which it is possible to transform into a circle a figure that may be described as an *irregular star*, composed of any number of rays of any lengths and at any angles, such as shown in fig. 2. Perhaps the most useful example of this would be an " aeroplane with dihedral angle ".

2. Since the purpose of this note is chiefly to put the transformations on record, there is no need to explain in detail how they were found. It need only be said that there are two primitive forms from which they were

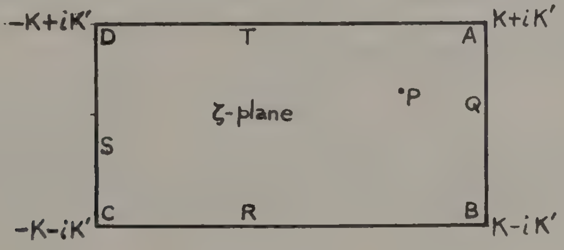
---

* *Aer. Res. Co., Rep. & Mem.* No. 2280 (Oct. 1945).

generalized which may be called the H and the X, from the shapes of those letters. The H was the subject of the previous note, though the notation has been improved from that used there. The interesting case is where the cross-bar is not exactly half way up it. By bisecting the H vertically the Schwartzian method can be used, and the most important point that emerges is that if it is to fit together properly, the elliptic integrals of the third kind which arise in the integration must vanish. This is then seen to be a general principle, because such an integral has a logarithmic singularity, whereas the only admissible singularity is a simple pole to represent the point at infinity. This condition very much simplifies all the work. Other problems involving four right-angled turns can also be solved by elliptic integrals, but in general they involve integrals of the third kind.

The X, or cross, does not lend itself directly to the Schwartzian method, though a symmetrical cross can be discussed in that way by twisting it through 45°. However, as will appear, it is very easy to guess the transformation directly, and there is therefore no need to explain this process.

Fig. 3.



3. In order to follow the various diagrams it will be well to explain their lettering. The four right-angled turns are lettered ABCD taken counter-clockwise in the z-plane. Since the outer region in z is to be transformed into the inside of a rectangle in ζ, these points go clockwise in ζ as shown in fig. 3. The corners of the rectangle are the points $\pm K \pm iK'$ in the ζ-plane. Any point on AB is labelled Q, and similarly for the other sides by R, S and T. In particular these letters will denote the turning points of the fins; if there are more than one on a side of the rectangle, they are given subscripts, and the subscript 0 will be used in the particular case where the mid-point of the side is referred to. One other place of importance in the ζ-rectangle is the pole P, which corresponds in the z-plane to the point at infinity.

4. The transformations all involve expressing z in terms of elliptic functions of ζ. In addition to the ordinary functions, we shall have elliptic integrals of the second kind, and the most convenient one to use is the one

that vanishes at both 0 and K. In the ordinary notation this is written with a term in E/K, but in the present work E will never occur by itself, so we shall adopt a single symbol and write

$$G=E/K. \qquad (4.1)$$

Then take

$$Z(\zeta)=\int_0^\zeta \mathrm{dn}^2\, u\, du-G\zeta. \qquad (4.2)$$

This function, like $\mathrm{sn}\,\zeta$ etc., is infinite at the points $\pm iK'$, and this is inconvenient. It is quite easy to deal with the infinity of $\mathrm{sn}\,\zeta$, etc., but it is better to eliminate it for Z by considering the function

$$Z(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta} \qquad (4.3)$$

instead. The second term destroys the infinities at $\pm iK'$, but makes a pole at the origin. It is then comparatively easy to shift this pole to any other point P.

To work out the values in z corresponding to the rectangle in ζ it is convenient to have a table, for though the values of the functions are all given by well known formulæ, there are a good many such formulæ, and their signs are rather confusing. Table I. gives the expressions used; the modulus is taken as k for sn, cn, dn, Z, but as k', its complement, when they are accented, as in sn′, cn′, dn′, Z′. The last row gives the ordinary Jacobian transformation.

Table I.

	ζ	$\mathrm{sn}\,\zeta$	$\mathrm{cn}\,\zeta$	$\mathrm{dn}\,\zeta$	$Z(\zeta)+\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta/\mathrm{sn}\,\zeta$
Q	$K+i\eta$	$\dfrac{1}{\mathrm{dn}'\eta}$	$-\dfrac{ik'\mathrm{sn}'\eta}{\mathrm{dn}'\eta}$	$\dfrac{k'\mathrm{cn}'\eta}{\mathrm{dn}'\eta}$	$-i\left\{Z'(\eta)+\dfrac{\eta\pi}{2KK'}\right\}$
R	$-iK'+\xi$	$\dfrac{1}{k\,\mathrm{sn}\,\xi}$	$\dfrac{i\,\mathrm{dn}\,\xi}{k\,\mathrm{sn}\,\xi}$	$\dfrac{i\,\mathrm{cn}\,\xi}{\mathrm{sn}\,\xi}$	$Z(\xi)+\dfrac{i\pi}{2K}$
S	$-K+i\eta$	$-\dfrac{1}{\mathrm{dn}'\eta}$	$\dfrac{ik'\mathrm{sn}'\eta}{\mathrm{dn}'\eta}$	$\dfrac{k'\mathrm{cn}'\eta}{\mathrm{dn}'\eta}$	$-i\left\{Z'(\eta)+\dfrac{\eta\pi}{2KK'}\right\}$
T	$iK'+\xi$	$\dfrac{1}{k\,\mathrm{sn}\,\xi}$	$-\dfrac{i\,\mathrm{dn}\,\xi}{k\,\mathrm{sn}\,\xi}$	$-\dfrac{i\,\mathrm{cn}\,\xi}{\mathrm{sn}\,\xi}$	$Z(\xi)-\dfrac{i\pi}{2K}$
	$i\tau$	$\dfrac{i\,\mathrm{sn}'\tau}{\mathrm{cn}'\tau}$	$\dfrac{1}{\mathrm{cn}'\tau}$	$\dfrac{\mathrm{dn}'\tau}{\mathrm{cn}'\tau}$	$-i\left\{Z'(\tau)+\dfrac{\mathrm{dn}'\tau\,\mathrm{cn}'\tau}{\mathrm{sn}'\tau}+\dfrac{\pi\tau}{2KK'}\right\}$

The following points may be noted. On the boundary $\mathrm{sn}\,\zeta$ is everywhere real, being always greater than 1 and positive on the right-hand side, negative on the left. $\mathrm{cn}\,\zeta$ is everywhere purely imaginary on the boundary, and its real part is everywhere positive inside. $\mathrm{dn}\,\zeta$ is real on Q and S, imaginary on R and T, while its real part is everywhere positive inside.

The function in the last column is imaginary on Q and S, vanishing at the mid-points, while on R and T it has constant imaginary parts equal and opposite in sign.

5. The transformation yielding the symmetrical cross is

$$z=\frac{\operatorname{dn}\zeta}{\operatorname{sn}\zeta}. \qquad (5.1)$$

This is easily generalized to an irregular cross as

$$z=\frac{\operatorname{dn}\zeta}{\operatorname{sn}\zeta-f-ig\operatorname{cn}\zeta}. \qquad (5.2)$$

where f and g are real, since the denominator is then real all round the contour. For the pole to fall inside the rectangle it is necessary that

$$f^2+g^2<1. \qquad (5.3)$$

The transformation yielding the H with unequal ends is

$$z=\mathrm{Z}(\zeta)+\frac{\operatorname{dn}\zeta\operatorname{cn}\zeta}{\operatorname{sn}\zeta-f} \qquad (5.4)$$

A corresponding generalization for H can be made, and combining the two formulae the final result is

$$\frac{z}{l}=\mathrm{Z}(\zeta)+\frac{\operatorname{dn}\zeta(\operatorname{cn}\zeta+ig\operatorname{sn}\zeta+ih)}{\operatorname{sn}\zeta-f-ig\operatorname{cn}\zeta}+m\zeta, \qquad (5.5)$$

where all the parameters l, k, f, g, h, m are real.

The first two terms are a generalization of those given in the last column of Table I. By adding and subtracting $\operatorname{dn}\zeta\operatorname{cn}\zeta/\operatorname{sn}\zeta$ in (5.5) it is easy to see that the second term still destroys the infinity of Z at R_0 and T_0, and now yields a single pole determined by the denominator of the second term. The general formulae corresponding to the last column of Table I. are not written out here, since they are rather complicated, though they may be worked out easily by the use of Table I.

The equation (5.5) has six parameters l, k, f, g, h and m, and by suitable choices these can be adapted to give the six dimensions of the finned rectangle, that is, its two sides and the lengths (and directions) of the four fins.

The turning points of the fins are given by equating to zero

$$\begin{aligned}\frac{1}{l}\frac{dz}{d\zeta}=m-\mathrm{G}+\{&-(1-f^2)+\operatorname{sn}^2\zeta(1-k^2f^2+k'^2g^2)+\operatorname{sn}\zeta k'^2(-f+gh)\\&+i\operatorname{cn}\zeta(-h+fg)+i\operatorname{sn}\zeta\operatorname{cn}\zeta[-g(1+k'^2)+fhk^2]\}{,}\,(\operatorname{sn}\zeta-f-ig\operatorname{cn}\zeta)^2,\end{aligned} \qquad (5.6)$$

which is a quartic in $\operatorname{sn}\zeta$.

Writing $p=\sqrt{(1-f^2-g^2)}$ the pole P is at ζ_0, where

$$\operatorname{sn}\zeta_0=\frac{f+igp}{1-g^2},\quad \operatorname{cn}\zeta_0=\frac{p-ifg}{1-g^2}. \qquad (5.7)$$

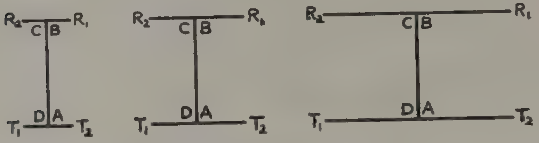
6. We now describe some of the shapes given by specializing the parameters. The diagrams are only sketches, in that when it is asserted that one parameter is changing no attention has been given to keeping the others strictly constant. To draw them accurately would involve much work and give little further information.

Take first the shapes that are symmetrical about two central lines at right angles. For these f, g and h vanish.

The first to be considered is

$$z=\mathrm{Z}(\zeta)+\frac{\operatorname{dn}\zeta \operatorname{cn}\zeta}{\operatorname{sn}\zeta}. \qquad (6.1)$$

Fig. 4



As k increases, this gives the sequence of fig. 4. The turning points are the solutions of

$$\frac{\operatorname{cn}^2\zeta}{\operatorname{sn}^2\zeta}=-\mathrm{G} \qquad (6.2)$$

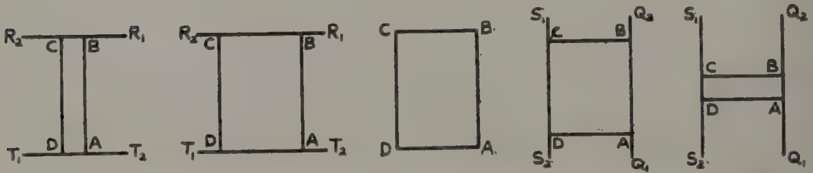
or transforming to the lines R or T by Table I.,

$$\operatorname{dn}^2\xi=\mathrm{G} \qquad (6.3)$$

With this value of ξ,

$$z=\pm\mathrm{Z}(\xi)\pm\frac{i\pi}{2\mathrm{K}} \qquad (6.4)$$

Fig. 5.



at the ends of the fins. When k is not large ξ is somewhere near $\frac{1}{2}$K, but tables of elliptic functions must be used to evaluate it accurately. The value of k must be very close indeed to 1 to get an H with long arms and a short cross-bar.

Next take

$$z=\mathrm{Z}(\zeta)+\frac{\operatorname{dn}\zeta \operatorname{cn}\zeta}{\operatorname{sn}\zeta}+m\zeta \qquad (6.5)$$

and consider how it behaves as m changes. The effect of m is to thicken up the rectangle as in fig. 5. It must be positive, for a negative m would cause AB to overlap CD the wrong way round.

The rectangle has no fins when $m=G-k'^2$ after which the fins switch through a right angle. At $m=\pi/2KK'$, BC coincides with AD, and the sequence stops. The second half of the sequence can be directly derived from the first by the Jacobian transformation.

For the special case $k=1/\sqrt{2}$, $m=G-\frac{1}{2}=\pi/4K^2$ the shape is a square, and the Jacobian transformation shows its tetragonal symmetry.

7. If now a term in h is included, the two lines of symmetry are destroyed, but the centre is still a centre of symmetry.

First take

$$z=Z(\zeta)+\frac{\operatorname{dn}\zeta\operatorname{cn}\zeta}{\operatorname{sn}\zeta}+\frac{ih\operatorname{dn}\zeta}{\operatorname{sn}\zeta}. \quad \dots \dots \quad (7.1)$$

As h increases from zero the fins alter as shown in fig. 6. When $h=(G-k'^2)/kk'$ one pair of fins disappears. As h grows to infinity the side-arms close down, leading to a pure cross when h becomes infinite. If a term in m is added, this merely thickens up the rectangle.

Fig. 6.

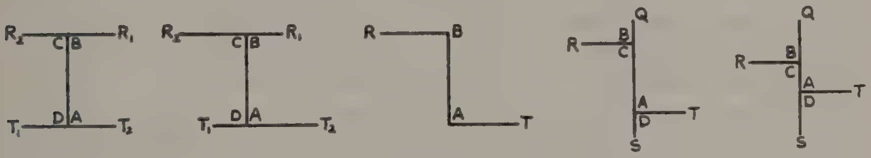
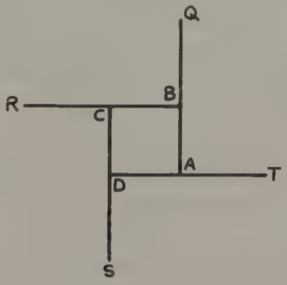


Fig. 7.



Attention may be drawn to the "windmill", fig. 7:

$$z=Z(\zeta)+\frac{\operatorname{dn}\zeta\operatorname{cn}\zeta}{\operatorname{sn}\zeta}+\frac{\pi}{4K^2}\zeta+\frac{ih\operatorname{dn}\zeta}{\operatorname{sn}\zeta}\left(k=\frac{1}{\sqrt{2}}\right),$$

which is tetragonally symmetrical. For those familiar with the peculiarities of the hydrodynamic fluid it is hardly necessary to say that no matter how strongly the wind blows, or in what direction, the windmill experiences no resultant couple at all!

8. Next consider cases where P is no longer at the origin. Take

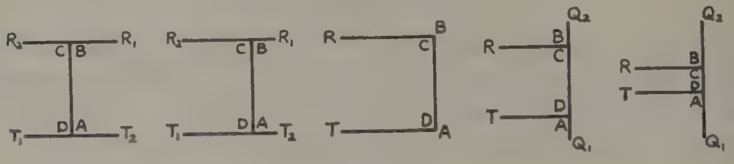
$$z=Z(\zeta)+\frac{\operatorname{dn}\zeta\operatorname{cn}\zeta}{\operatorname{sn}\zeta-f}. \quad \dots \dots \quad (8.1)$$

For growing f the family of shapes is then as in fig. 8. The transitional case with only two fins occurs for

$$f=(G-k'^2)/kG. \qquad (8.2)$$

If a term in m is added, the rectangle is merely thickened as before. In both these cases a quadratic in $\operatorname{sn}\zeta$ determines the values of ξ or η at the turning points, the two pairs of fins being associated with the two signs of $i\operatorname{cn}\zeta$.

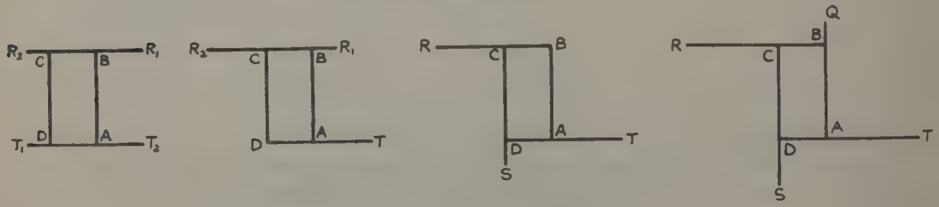
Fig. 8.



If further a term in h is added, the matter becomes more complicated because the equation determining the turning points is an irresoluble quartic, so that numerical solutions would have to be done. Fig. 9 sketches the shapes as h increases from zero to large values.

9. The case where f vanishes, but not g, can be reduced by the Jacobian transformation to the case of g vanishing and not f. When none of the constants vanish the matter becomes very complicated. If numerical values are assigned to all the constants, it is comparatively easy to find the shape. In doing so it is easiest not to use (5.6) in order to find the turning points, but instead merely to estimate the maxima from (5.5) by taking out a few values in their neighbourhoods. Thus fig. 1 is a somewhat exaggerated sketch given by the values

Fig. 9.



$$l=1, \quad k=0{\cdot}5, \quad f=0{\cdot}4, \quad g=0{\cdot}3, \quad h=0{\cdot}1, \quad m=0{\cdot}2.$$

The corners are at $\pm 0{\cdot}337 \pm i0{\cdot}501$ and the fins at

$$Q_1=0{\cdot}337-i\,0{\cdot}565 \qquad R=-0{\cdot}339+i0{\cdot}501$$

$$Q_2=0{\cdot}337+i\,0{\cdot}866 \qquad T=-0{\cdot}386-i0{\cdot}501$$

The converse problem of finding the parameters for a given shape is much more troublesome. I explored the method of solution by taking the comparatively simple problem of finding the parameters for a cross with arms of lengths 1, 2, 3, 4. This has the transformation

$$z=\frac{ih\,\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta-f-ig\,\mathrm{cn}\,\zeta}.$$

Starting values were guessed for k, f, g by considering the values for the half-symmetrical cases, where one pair of opposite arms are equal while the other pair have the required ratio, and with these the arms were calculated. It was also possible to set down an equation for the variation of the parameters, so that the effect of δf, etc., on all four arms could be seen; most of the quantities for these variations occur already in the main calculation. Three stages led to the values

$$h=2{\cdot}835,\quad k=0{\cdot}853,\quad f=0{\cdot}489,\quad g=0{\cdot}294,$$

which are correct to about one figure in the last place. The ends of the arms are then at 4, $3i$, -2, $-i$.

A similar method should work in the general case. The corners of the rectangle immediately give two relations between l, m and k. For the rest the method of trial must be used, but there is one control on the starting values of f, etc., which would help considerably. The value of $dz/d\zeta$ takes a fairly simple form at each of the corners, by the consideration that in differentiating (5.5) $\mathrm{dn}\,\zeta$ vanishes at the corners. Then according as $dz/d\zeta$ changes sign from one corner to the next or not, there will be one fin on that side or else either two or none. In this way some guidance can be given as to appropriate starting values for k, f, g, h, and the variation of parameters can then be used to get improvement. It may be noted that variation of k is a troublesome process, even if $\mathrm{sn}\,\zeta$ instead of ζ is used as the independent variable, so that for this one parameter it might be simplest in practice to omit this variation, and merely to find the shape twice over for two different close values of k. However, there is no doubt that the solution would be arduous, and it seemed hardly worth while to make up a purely artificial problem in order to solve it.

10. It will often be convenient to apply a second transformation, converting the ζ-rectangle into the exterior of a circle in such a manner that the point at infinity becomes the old point at infinity of the z-plane.

A simple case will show the method. The symmetrical cross is given by

$$z=\frac{\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta}. \qquad (10.1)$$

Write

$$z_0=\frac{1+\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta}. \qquad (10.2)$$

Then the point at infinity in z_0 is the same as in z. On the sides of the rectangle $\mathrm{sn}\,\zeta$ is real and $\mathrm{cn}\,\zeta$ imaginary, so that $|z_0|^2=1$, and so the rectangle becomes a circle of unit radius in the z_0-plane.

In the general case the transformation is

$$z_0=\frac{1}{\sqrt{(1-g^2)}}\,\frac{(1-g^2)-(f-igp)\,\mathrm{sn}\,\zeta+(p+ifg)\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta-f-ig\,\mathrm{cn}\,\zeta}, \quad . \quad (10.3)$$

where $p=\sqrt{(1-f^2-g^2)}$.

11. If it is desired to expose the finned rectangle to a "wind", it is necessary to form a function that has a pole at P, and that is real on the rectangle. Such a function is

$$w=\frac{\mathrm{A}\,\mathrm{sn}\,\zeta+i\mathrm{B}\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta-f-ig\,\mathrm{cn}\,\zeta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.1)$$

in which A and B are real. From this the components of the wind at infinity are given by the real and imaginary parts of

$$\left(\frac{dw}{dz}\right)_\infty=\frac{\mathrm{A}\,\mathrm{sn}\,\zeta_0+i\mathrm{B}\,\mathrm{cn}\,\zeta_0}{\mathrm{dn}\,\zeta_0(\mathrm{cn}\,\zeta_0+ig\,\mathrm{sn}\,\zeta_0+ih)}, \quad . \quad . \quad . \quad . \quad (11.2)$$

where ζ_0 is given by (5.7), $\mathrm{dn}\,\zeta_0$ having its real part positive.

An irrotational circulation round the rectangle is given by

$$w=i\mathrm{C}\ln z_0,$$

where z_0 is given by (10.3), since $\ln z_0$ is pure imaginary on the rectangle.

12. If it is desired to transform the finned rectangle into a circle, the intermediate use of the rectangle is in general indispensable, but there is an exception in the case of the cross, since the transformation then involves no elliptic integrals of the second kind. For example, between (10.1) and (10.2) ζ may be eliminated. Taking the special case of the equal armed cross with $k=1/\sqrt{2}$, the eliminant is

$$z=\tfrac{1}{2}(z_0^2+z_0^{-2})^{\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.1)$$

This may be compared with $z=\frac{1}{2}(z_0+z_0^{-1})$, which transforms the circle into a line bounded at both ends. This suggests immediately a generalization to a star of six, eight, etc., equal rays, but a very much wider generalization than this can be made.

Take

$$z=lz_0^{-1}(z_0-e^{i\alpha_1})^{\lambda_1}(z_0-e^{i\alpha_2})^{\lambda_2}\ldots(z_0-e^{i\alpha_n})^{\lambda_n}. \quad . \quad . \quad . \quad (12.2)$$

where

$$\lambda_1+\lambda_2+\ldots+\lambda_n=2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.3)$$

To study the transformation put $z=e^{i\theta}$, and consider the part where $\alpha_1<\theta<\alpha_2$. For this part, using (12.3),

$$z=4l\exp\{i\tfrac{1}{2}\Sigma_1^n\alpha_s\lambda_s+i\pi(\lambda_1-1)\}\sin^{\lambda_1}\tfrac{1}{2}(\theta-\alpha_1)\sin^{\lambda_2}\tfrac{1}{2}(\alpha_2-\theta)\ldots\sin^{\lambda_n}\tfrac{1}{2}(\alpha_n-\theta), \quad . \quad . \quad . \quad (12.4)$$

so that z traces a ray in the direction $\frac{1}{2}\Sigma_1^n\alpha_s\lambda_s+\pi(\lambda_1-1)$, vanishing at $\theta=\alpha_1$ and α_2, and so having a maximum in between. When $\alpha_2<\theta<\alpha_3$ there will be another ray in the direction $\frac{1}{2}\Sigma_1^n\alpha_s\lambda_s+\pi(\lambda_2+\lambda_1-1)$. Thus there are rays of various lengths at successive angles $\pi\lambda_1, \pi\lambda_2, \ldots \pi\lambda_n$ to one another. By suitable choice of l, $n-1$ of the λ's and the n α's any arbitrary irregular star of the type shown in fig. 2 can be transformed into a circle.

In finding this transformation the λ's can be given at once, but the α's would have to be found by trial. It may be noted that if two adjacent α's are taken equal, one of the rays disappears, and so it may be concluded that if two α's are close together, the ray between them will be short. We will be content with two examples.

Take

$$z=lz_0^{-1}(z_0-1)^{2\beta}(z_0+1)^{2-2\beta}. \qquad (12.5)$$

This represents a dihedral with equal arms at angle $2\pi\beta$ apart, as in fig. 10. The maxima occur at $z_0=\exp(\pm i\theta_1)$, where $\cos\theta_1=1-2\beta$. The rays are of length $4l\beta^\beta(1-\beta)^{1-\beta}$.

Fig. 10.

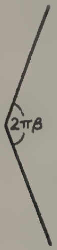
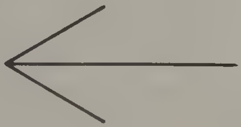


Fig. 11.



As a second example I arbitrarily chose the problem of an arrow-head, in which the barbs, at 30° to the shaft, were only half as long, as in fig. 11. This is represented by the transformation

$$z=z_0^{-1}(z_0-e^{i\alpha})^{1/6}(z_0+1)^{5/3}(z_0-e^{-i\alpha})^{1/6}, \qquad (12.6)$$

and without much trouble it was found that $\alpha=71°\,46'$. The shaft is then of length 3·348 and the barbs 1·673.

II. *The Propagation of Sound Waves in an Open-ended Channel.*

By W. Chester,
Department of Mathematics, Leeds University *.

[Received August 22, 1949.]

Summary.

An exact solution is obtained for the problem of a harmonic wave of sound approaching the open end of a two-dimensional channel, and suffering reflection and transmission. Within a certain wavelength

---

* Communicated by Professor S. Goldstein, F.R.S.

range, the wave returning along the channel becomes sensibly plane at large distances from the mouth, the reflection coefficient being $\exp[-2\pi b/\lambda]$, where $2b$ is the width of the channel and λ is the wavelength. Explicit formulæ are also obtained for the change of phase in the returning wave, and for the transmitted energy.

It is also shown that the solution can be used to deduce the energy absorbed from a plane wave external to the channel and approaching from an arbitrary direction.

INTRODUCTION.

WHEN a plane harmonic wave of sound is incident upon a two-dimensional channel, formed by two semi-infinite parallel planes, diffraction effects will occur at the open end. Ultimately, a periodically steady state is reached, and the object of the present paper is to determine the asymptotic form of the resulting periodically steady potential field, assuming that the wavelength lies within a certain range so that only the dominant mode is propagated inside the channel. There is no theoretical difficulty in extending the analysis to include a wider wavelength range, but, from a physical view-point, the present range is considered sufficient.

Problems of this type were first considered by Rayleigh using approximate methods. For the particular case of an incident wave originating inside a circular pipe, Rayleigh (1940) gave an estimate of the end correction when the wavelength was large, and also deduced that it increased indefinitely with the wavelength for a two-dimensional channel (Rayleigh 1904). (The end correction is the correction to the length of the pipe in order that the open end may be treated as a loop).

More recently, Levine and Schwinger (1948) gave an exact solution for the circular pipe, valid for a wave-length range in which only dominant mode propagation occurs. The present investigation gives the analogous solution for the two-dimensional channel, and is inspired by the analysis of the above authors.

A detailed investigation is given for an incident wave approaching the open end from inside the channel. The whole of the asymptotic behaviour of the wave, both inside and outside the channel, can be described in terms of the discontinuity in potential across the channel walls, and the problem consists, essentially, of deriving an integral equation for this quantity which is susceptible to rigorous, analytical solution. Rayleigh's conclusion is verified and, in general, the final results assume a simple form. The end correction and transmitted energy are exhibited graphically in figures 5 and 6.

It is also shown that, having obtained the solution for the above problem, the energy absorbed by the channel from a plane wave approaching the open end from an arbitrary direction can then be deduced immediately with the aid of a simple reciprocity relation which correlates the radiating and absorbing properties of the channel.

Asymptotic Behaviour.

Since the analysis is eventually confined to the behaviour at large distances from the mouth of the channel, we begin with some general remarks concerning the asymptotic form assumed by the potential both inside and outside the channel.

The origin of coordinates is taken at the mouth of the channel with the negative x-axis along the centre-line. The potential of the incident wave is denoted by $A \exp [ik(x-at)]$ or $\exp [-ik\{r \cos (\theta-\theta_0)-at\}]$ according as it originates inside the channel or approaches from a direction θ_0, and the final potentials resulting from these two incident waves are denoted by ϕ_a, ϕ_b respectively.

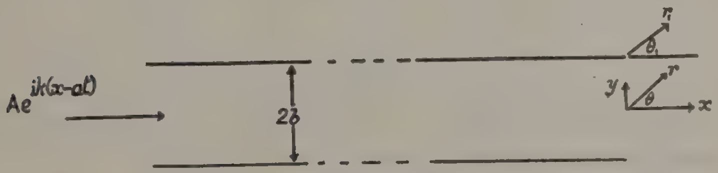
Inside the channel, the most general solution of the wave equation $(\nabla^2+k^2)\phi=0$, expanded as a Fourier series in y and satisfying the boundary condition of zero normal derivative at $y=\pm b$, is of the form

$$\begin{aligned}\phi=A \exp [ikx]+B \exp [-ikx]+ \sum_{n=1}^{\infty} (A_n \exp [x(q^2-k^2)^{\frac{1}{2}}]\\ +B_n \exp [-x(q^2-k^2)^{\frac{1}{2}}] \cos qy+ \sum_{n=0}^{\infty} (A'_n \exp [x(s^2-k^2)^{\frac{1}{2}}]\\ +B'_n \exp [-x(s^2-k^2)^{\frac{1}{2}}]) \sin sy, \quad x<0, \quad \ldots \ldots \quad (1)\end{aligned}$$

where

$$q=n\pi/b, \quad s=(2n+1)\pi/2b.$$

Fig. 1.



(The factor $\exp [ikat]$ here, and in the subsequent analysis, is suppressed).

The potential ϕ_a is symmetrical about the centre line of the channel, so that, in this case, $A'_n=B'_n=0$. It is now clear from equation (1), that the condition for dominant mode propagation is $bk<\pi$. All the coefficients B_n must then be zero if ϕ_a is to remain finite for large negative x, and so we deduce that, as $x \to -\infty$ inside the channel,

$$\phi_a \sim A \exp [ikx]+B \exp [-ikx]. \quad \ldots \ldots \quad (2)$$

Similar remarks apply to ϕ_b except that A'_n is no longer zero. This reduces the range of k to $kb<\pi/2$; then

$$\phi_b \sim C \exp [-ikx], \quad \ldots \ldots \quad (3)$$

since there is now no incident wave inside the channel.

In practice, the above restrictions will usually be satisfied. The condition $bk<\pi/2$, for example, is equivalent to $b/\lambda<\frac{1}{4}$, and for a typical sound frequency of 270 oscillations per second, the wavelength is four feet.

Outside the channel, ϕ_a will behave, asymptotically, like a source situated at the origin whose amplitude varies with direction. Thus

$$\phi_a \sim f(\theta)\frac{\exp[ikr]}{r^{\frac12}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The behaviour of ϕ_b is, however, somewhat more complex. If $\frac{\pi}{2}<\theta_0<\pi$, it must take account of the incident wave which persists in the region $-(\pi-\theta_0)<\theta_1<\pi$, and a wave reflected from the upper wall in the region $(\pi-\theta_0)<\theta_1<\pi$, where θ_1 is the polar coordinate referred to the upper lip of the channel (fig. 1). It follows that the correction term due to diffraction at the mouth of the channel cannot tend to zero at infinity along $\theta_1=\pm(\pi-\theta_0)$ since the continuity of the potential must be maintained in the transition from one region to another.

To overcome this difficulty, appeal is made to the known solution for the diffraction of a harmonic wave by a semi-infinite wall. This problem was first solved by Sommerfeld (1894) and later discussed by Carslaw (1898). In terms of coordinates (r_1, θ_1) about an origin at the edge of the wall (fig. 1), a wave $\exp[-ikr_1\cos(\theta_1-\theta_0)]$ is diffracted to give a potential,

$$\left(\frac{k}{\pi}\right)^{\frac12}\exp[-i\pi/4]\left\{\exp[-ikr_1\cos(\theta_1-\theta_0)]\int_{-\infty}^{(2r_1)^{\frac12}\cos\frac12(\theta_1-\theta_0)}\exp[ik\lambda^2]\,d\lambda\right.$$
$$\left.+\exp[-ikr_1\cos(\theta_1+\theta_0)]\int_{(2r_1)^{\frac12}\cos\frac12(\theta_1+\theta_0)}^{\infty}\exp[ik\lambda^2]\,d\lambda.\right\} \quad . \quad . \quad (5)$$

It is easily verified that the normal derivative of this expression is zero along the wall $(\theta_1=\pm\pi)$, and since

$$\left.\begin{aligned}\int_{-\infty}^{(2r_1)^{\frac12}\cos\frac12\theta}\exp[ik\lambda^2]\,d\lambda &\sim \left(\frac{\pi}{k}\right)^{\frac12}\exp[i\pi/4]+\frac{\exp[ikr_1(1+\cos\theta)]}{2^{\frac32}ik^{\frac12}r^{\frac12}\cos\frac12\theta}\quad 0<\theta<\pi-\epsilon,\\ &\sim \frac{\exp[ikr_1(1+\cos\theta)]}{2^{\frac32}ik^{\frac12}r^{\frac12}\cos\frac12\theta}\quad \pi+\epsilon<\theta<\pi,\end{aligned}\right\} \quad . \quad . \quad . \quad (6)$$

it follows that the above potential includes the plane wave contributions in the relevant intervals, together with a term which behaves like $\exp[ikr]/r^{\frac12}$ times a function of θ_1, except along the lines $\theta_1=\pm(\pi-\theta_0)$. This last peculiarity of the solution is essential if the potential is to remain continuous in the transition from one region to another.

In the present problem there will be further diffraction effects in addition to those given by (5), due to the presence of the second diffracting wall of the channel. Their contribution to the potential must, however, take the form $g(\theta_1)\exp[ikr_1]/r^{\frac{1}{2}}$ for large r, and the asymptotic behaviour of ϕ_b is, accordingly,

$$\phi_b \sim \left(\frac{k}{\pi}\right)^{\frac{1}{2}} \exp[-i\pi/4-ikb\sin\theta_0]\Big\{\exp[-ikr_1\cos(\theta_1-\theta_0)] \times\int_{-\infty}^{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}(\theta_1-\theta_0)} \exp[ik\lambda^2]\,d\lambda+\exp[-ikr_1\cos(\theta_1+\theta_0)] \times\int_{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}(\theta_1+\theta_0)}^{\infty} \exp[ik\lambda^2]\,d\lambda\Big\}+g(\theta_1)\frac{ikr_1}{r_1^{\frac{1}{2}}} \quad . \; . \; . \; . \; . \; . \quad (7)$$

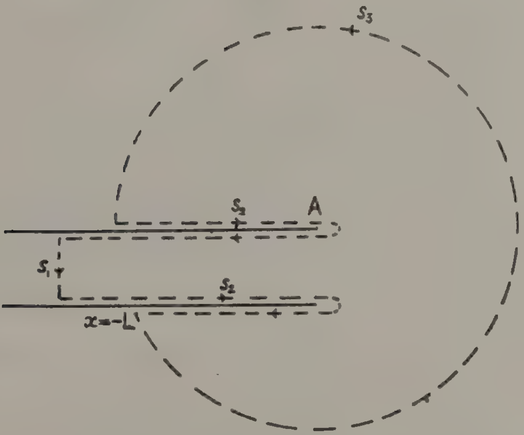
outside the channel.

The factor $\exp[-ikb\sin\theta_0]$ is due to the change in origin. An incident wave $\exp[-ikr\cos(\theta-\theta_0)]$ becomes

$$\exp[-ikr_1\cos(\theta_1-\theta_0)-ikb\sin\theta_0]$$

when referred to coordinates (r_1, θ_1) at the upper lip of the channel.

Fig. 2.



Strictly, equation (7) has been derived for $\pi/2<\theta_0<\pi$ since, for $0<\theta_0<\pi/2$, the boundary for the incident wave term in ϕ_b is a line parallel to $\theta_1=-(\pi-\theta_0)$ and passing through the lower lip of the channel. It can, however, be shown that the above expression does represent the *asymptotic* behaviour of ϕ_b for $0<\theta_0<\pi$.

RECIPROCITY RELATION.

In this section we show that the radiation characteristics of the channel can be used to determine the energy absorbed from a wave originating outside the channel.

To obtain this reciprocity relation, we apply Green's Theorem,

$$\iint_S \{\phi_a\nabla^2\phi_b-\phi_b\nabla^2\phi_a\}dS=\int_s\left\{\phi_a\frac{\partial\phi_b}{\partial n}-\phi_b\frac{\partial\phi_a}{\partial n}\right\}ds \quad . \; . \quad (8)$$

to the two potentials ϕ_a, ϕ_b, the domain of integration being that bounded by the curves s_1, s_2, s_3 (fig. 2), and n being the outward normal. The portion of the bounding curve denoted by s_2 is taken along the upper and lower sides of each wall of the channel, while s_3 is a large circle, centre A, the origin of the coordinates (r_1, θ_1). Both s_1 and s_3 are considered to be at a sufficiently large distance from the origin to enable the asymptotic relations for ϕ_a, ϕ_b to be used.

Since both ϕ_a and ϕ_b satisfy the equation

$$(\nabla^2+k^2)\phi=0$$

in the region considered, and since, further, they have zero normal derivatives at the walls, equation (8) reduces to

$$\int_{s_1}\left\{\phi_a\frac{\partial\phi_b}{\partial n}-\phi_b\frac{\partial\phi_a}{\partial n}\right\}ds+\int_{s_3}\left\{\phi_a\frac{\partial\phi_b}{\partial n}-\phi_b\frac{\partial\phi_a}{\partial n}\right\}ds=0. \quad . \quad . \quad (9)$$

Substituting from equations (2) and (3), the first integral becomes

$$-\int_{-b}^{b}\Big[(\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx])\frac{\partial}{\partial x}(\mathrm{C}\exp[-ikx])-\mathrm{C}\exp[-ikx]$$
$$\times\frac{\partial}{\partial x}(\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx])\Big]_{x=-\mathrm{L}}dy=4ibk\mathrm{AC}. \quad . \quad . \quad . \quad (10)$$

In evaluating the second integral in equation (9) we note, from equation (7), that the behaviour of ϕ_b on s_3 is equivalent to that of the terms $\exp[-ikr_1\cos(\theta_1-\theta_0)-ikb\sin\theta_0]$ and $\exp[-ikr_1\cos(\theta_1+\theta_0)-ikb\sin\theta_0]$ in the intervals $-(\pi-\theta_0)+\epsilon<\theta_1<\pi$ and $(\pi-\theta_0)+\epsilon<\theta_1<\pi$ respectively together with a term which behaves like $h(\theta_1)\cdot\exp[ikr_1]/r_1^{\frac{1}{2}}$, except in the intervals $|\theta_1\pm(\pi-\theta_0)|<\epsilon$. It can be shown that the contribution to the integral along s_3 from the two small intervals $|\theta_1\pm(\pi-\theta_0)|<\epsilon$ is $0(\epsilon)$ in the limit $r_1\to\infty$, while the contribution from the term $h(\theta_1)\exp[ikr_1]/r_1^{\frac{1}{2}}$ is

$$\int\Big\{f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\frac{\partial}{\partial r_1}\left(h(\theta_1)\frac{\exp[ikr_1]}{r_1^{\frac{1}{2}}}\right)$$
$$-h(\theta_1)\frac{\exp[ikr_1]}{r_1^{\frac{1}{2}}}\frac{\partial}{\partial r_1}\left(f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\right)\Big\}r_1\,d\theta_1=0, \quad . \quad . \quad (11)$$

integrated over any range of θ_1, since the integrand is identically zero.

Again, a term $\exp[-ikr_1\cos\theta_1]$ (say) in ϕ_b will give a contribution

$$\int\Big\{f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\frac{\partial}{\partial r_1}(\exp[-ikr_1\cos\theta_1])-\exp[ikr_1\cos\theta_1]$$
$$\times\frac{\partial}{\partial r_1}\left(f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\right)\Big\}r_1\,d\theta_1$$

$$=-ikr_1^{\frac{1}{2}}\int f(\theta_1)\exp[ikb\sin\theta_1+ikr_1(1-\cos\theta_1)](1+\cos\theta_1)\,d\theta_1+0(r^{-\frac{1}{2}})$$

$$=2^{\frac{3}{2}}ik\int f(\theta_1)\exp[ikb\sin\theta_1]\cos\theta_1/2\,\frac{\partial}{\partial\theta}\left[\pm\int_{(2r_1)^{\frac{1}{2}}\left|\sin\frac{\theta_1}{2}\right|}^{\infty}\exp[ik\lambda^2]\,d\lambda\right]$$
$$\times d\theta_1+0(r^{-\frac{1}{2}}), \quad . \quad . \quad (12)$$

the plus or minus sign to be taken according as θ_1 is positive or negative.

A simple integration by parts now shows that this expression tends to zero as $r_1 \to \infty$ unless the range of integration includes the point $\theta_1 = 0$.

Thus, for $0 < \theta_0 < \pi$, the only contribution to the integral along s_3, in the limit as $r_1 \to \infty$, occurs from the term

$$\exp\,[-ikr_1 \cos(\theta_1 - \theta_0) - ikb \sin\theta_0],$$

in a small interval surrounding $\theta_1 = \theta_0$. The term

$$\exp\,[-ikr_1 \cos(\theta_1 + \theta_0) - ikb \sin\theta_0]$$

has no contribution, in the limit, since it occurs only for $(\pi - \theta_0) < \theta_1 < \pi$ which exludes the point $\theta_1 = -\theta_0$.

It follows that

$$\lim_{r_1 \to \infty} \int_{s_3} \left(\phi_a \frac{\partial \phi_b}{\partial n} - \phi_b \frac{\partial \phi_a}{\partial n} \right) ds$$

$$= - \lim_{r_1 \to \infty} ikr_1^{\frac{1}{2}} \int_{\theta_0 - \delta}^{\theta_0 + \delta} f(\theta_1) \exp\,[ikb(\sin\theta_1 - \sin\theta_0) + ikr_1\{1 - \cos(\theta_1 - \theta_0)\}]$$
$$\times (1 + \cos(\theta_1 - \theta_0))\, d\theta_1$$

$$= - \lim_{r_1 \to \infty} 2ikr_1^{\frac{1}{2}} f(\theta_0) \int_{-\delta}^{\delta} \exp\,[2ikr_1 \sin^2 \theta/2] \cos\theta/2\, d\theta$$

$$= -4ikf(\theta_0) \int_{-\infty}^{\infty} \exp[2ikz^2]\, dz$$

$$= -4ik \left(\frac{\pi}{2k} \right)^{\frac{1}{2}} f(\theta_0) \exp\,[i\pi/4]. \qquad (13)$$

Finally, equating (10) and (13), we have

$$C = \left(\frac{\pi}{2k} \right)^{\frac{1}{2}} \frac{\exp\,[i\pi/4] f(\theta_0)}{Ab}. \qquad (14)$$

The limiting case $\theta_0 = \pi$ is anomalous. There is now no reflected wave, but the term which takes account of this wave in equation (5) does not tend to zero for $\theta_0 = \pi$ and the whole expression for ϕ_b is asymptotic to $2 \exp\,[ikr_1 \cos\theta_1]$ for $0 < \theta_0 < \pi$. The asymptotic behaviour of ϕ_b for, an incident wave, $\exp\,[ikr_1 \cos\theta_1]$ is, in fact,

$$\phi_b \sim \left(\frac{k}{\pi} \right)^{\frac{1}{2}} \exp\,[-i\pi/4 + ikr_1 \cos\theta_1] \int_{-\infty}^{(2r_1)^{\frac{1}{2}} \sin\frac{1}{2}\theta} \exp\,[ik\lambda^2]\, d\lambda + g(\theta_1) \frac{\exp\,[ikr_1]}{r_1^{\frac{1}{2}}} \qquad (15)$$

which, for large r, behaves, as it should, like $\exp\,[ikr_1 \cos\theta_1]$ for $0 < \theta_1 < \pi$.

The analogous relation to (14) is then

$$C(\pi) = \left(\frac{\pi}{2k} \right)^{\frac{1}{2}} \frac{\exp\,[i\pi/4] f(\pi)}{2Ab}. \qquad (16)$$

Equations (14) and (16) enable us to deduce the energy absorbed from a wave approaching from any direction θ_0, in terms of the asymptotic behaviour of a wave originating inside the channel.

Expressions for the Reflected and Transmitted Waves.

The analysis is now confined to the asymptotic behaviour of the wave originating inside the channel which, as we shall see, can be described wholly in terms of the discontinuity in ϕ_a across the walls of the channel.

To prove this, we apply Green's Theorem to the potential ϕ_a, and a plane wave $\exp[-ikr\cos(\theta-\theta_0)]$. The domain of integration is that shown in fig. 2 except that the centre of the large circle is now taken to be 0, the origin of the (r, θ) coordinates.

The surface integral again vanishes and, using the asymptotic expressions for ϕ_a and the fact that its normal derivative vanishes along the walls of the channel, we find that

$$\begin{aligned}
&\int_{s_2}\left[f(\theta)\frac{\exp[ikr]}{r^{\frac{1}{2}}}\frac{\partial}{\partial r}(\exp[-ikr\cos(\theta-\theta_0)])-\exp[-ikr\cos(\theta-\theta_0)]\right.\\
&\quad\left.\times\frac{\partial}{\partial r}\left(f(\theta)\frac{\exp[ikr]}{r^{\frac{1}{2}}}\right)\right]ds\\
&\quad-\int_{-\mathrm{L}}^{0}[\phi_a]_{y=b}\frac{\partial}{\partial y}\exp[-ikx\cos\theta_0-iky\sin\theta_0]_{y=b}\,dx\\
&\quad-\int_{-\mathrm{L}}^{0}[\phi_a]_{y=-b}\frac{\partial}{\partial y}\exp[-ikx\cos\theta_0-iky\sin\theta_0]_{y=-b}\,dx\\
&\quad-\int_{s_1}\left\{(\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx])\frac{\partial}{\partial x}\exp[-ikx\cos\theta_0-iky\sin\theta_0]\right.\\
&\quad-\exp[-ikx\cos\theta_0-iky\sin\theta_0]\frac{\partial}{\partial x}(\mathrm{A}\exp[ikx]\\
&\quad\left.+\mathrm{B}\exp[-ikx])\right\}_{x=-\mathrm{L}}ds=0, \quad \ldots \ldots \quad (17)
\end{aligned}$$

where

$$[\phi_a]=\phi_a)_{y+0}-\phi_a)_{y-0} \quad \text{and} \quad [\phi_a]_{y=b}=-[\phi_a]_{y=-b}=\mathrm{H}(x) \text{ (say)}. \quad . \quad (18)$$

The first integral, by a similar argument to that used in the derivation of (13) is, in the limit as $r \to \infty$,

$$-2^{\frac{3}{2}}\pi^{\frac{1}{2}}ik^{\frac{1}{2}}f(\theta_0)\exp[i\pi/4]. \quad \ldots \ldots \quad (19)$$

Substituting in equation (17), and simplifying the remaining terms, we get

$$\begin{aligned}
\exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(\theta)=\lim_{\mathrm{L}\to\infty}&\left[k\sin\theta\sin(kb\sin\theta)\int_{-\mathrm{L}}^{0}\mathrm{H}(x)\right.\\
&\times\exp[-ikx\cos\theta]\,dx+\frac{i\sin(kb\sin\theta)}{\sin\theta}\left\{\mathrm{A}(1+\cos\theta)\exp[-ik\mathrm{L}(1-\cos\theta)]\right.\\
&\left.\left.-\mathrm{B}(1-\cos\theta)\exp[ik\mathrm{L}(1+\cos\theta)]\right\}\right] \quad \ldots \ldots \quad (20)\\
&=\mathrm{F}(k\cos\theta) \quad \text{(say)}. \quad \ldots \ldots \quad (21)
\end{aligned}$$

Since the asymptotic behaviour of $H(x)$ will be the same as that of ϕ_a inside the channel, reference to equation (2) shows that the oscillating terms on the right-hand side of equation (20) are cancelled by equal and opposite contributions from the integral, and the limit is uniquely defined.

It is convenient, at this point, to assume that k has a small positive imaginary part which will eventually be reduced to zero at an appropriate point in the analysis. Considered as a function of the complex variable ζ, $F(\zeta)$ then becomes, for $Im\zeta > Imk$,

$$F(\zeta)=(k^2-\zeta^2)^{\frac{1}{2}} \sin\{b(k^2-\zeta^2)^{\frac{1}{2}}\}H(\zeta), \quad . \quad . \quad . \quad . \quad (22)$$

where

$$H(\zeta)=\int_{-\infty}^{0} H(x) \exp[-i\zeta x]\, dx, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

and is the Fourier transform of a function equal to $H(x)$ when x is negative and zero for positive x. For economy of notation the same symbol is used for a function and its transform, the distinction being made by the argument of the function.

We may, therefore, regard (20) as the analytic continuation of $F(\zeta)$ in the interval $-Imk \leqslant Im\zeta \leqslant Imk$, so that, for $\zeta = k\cos\theta$

$$F(k\cos\theta)=\exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(\theta)=k\sin\theta\sin(kb\sin\theta)H(k\cos\theta). \qquad (24)$$

The asymptotic form of $H(x)$ given by equation (2), shows, with the help of (23), that $H(\zeta)$ has simple poles at $\zeta=\pm k$. Further, from (20), (22)

$$\left.\begin{aligned} \exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(0)&=2ikbA=-2bk\lim_{\zeta\to k}(\zeta-k)H(\zeta)=-2bk\operatorname*{Res}_{\zeta=k}H(\zeta),\\ \exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(\pi)&=-2ikbB=2bk\lim_{\zeta\to -k}(\zeta+k)H(\zeta)=2bk\operatorname*{Res}_{\zeta=-k}H(\zeta),\end{aligned}\right\} \quad . \quad . \quad . \quad (25)$$

and the reflection coefficient, R, becomes

$$R=\frac{B}{A}=\frac{\operatorname*{Res}_{\zeta=-k}H(\zeta)}{\operatorname*{Res}_{\zeta=k}H(\zeta)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

To describe the transmitted wave, we define the function

$$G(\theta)=\frac{2\pi|f(\theta)|^2}{2b(|A|^2-|B|^2)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where $|f(\theta)|^2$ measures the energy per unit angle radiated in a direction θ, and $2b(|A|^2-|B|^2)$, the total average energy transmitted from the mouth of the channel. The function $G(\theta)$ is therefore the ratio of the time average energy radiated in a direction θ compared with the average energy which would be radiated by a simple source at the origin of the same total strength.

Equations (24), (25) and (26) also show that

$$G(\theta)=\frac{2bk|f(\theta)|^2}{|f(0)|^2(1-|R|^2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (28)$$

$$=\frac{1}{2bk}\frac{\{k\sin\theta\sin(kb\sin\theta)\}^2|H(k\cos\theta)|^2}{|\operatorname*{Res}_{\zeta=k} H(\zeta)|^2-|\operatorname*{Res}_{\zeta=-k} H(\zeta)|^2}, \quad . \quad . \quad . \quad . \qquad (29)$$

which depends only on the function $H(\zeta)$.

Finally, in terms of $G(\theta)$, the amplitude of the absorbed wave, given by (14), satisfies the relation

$$|C|^2=G(\theta)/G(0). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (30)$$

The object of the rest of the paper is to obtain explicit expressions for the reflection coefficient R, and the function $G(\theta)$, which reduces the analysis, effectively, to the derivation of $H(\zeta)$, the transform of $H(x)$, since the relevant quantities depend on this function.

We show first that $H(x)$ satisfies an integral equation similar to the Wiener-Hopf type (Titchmarsh 1937). The usual method of solution of this equation is well adapted to the present problem since it is formulated in terms of the transform of $H(x)$, with which we are directly concerned.

THE INTEGRAL EQUATION FOR $H(x)$.

The boundary conditions at our disposal in deriving $H(x)$ are those of zero normal velocity at the walls of the channel, together with the known asymptotic forms of the potential at infinity. Such problems can be solved by expressing the potential in terms of an appropriate Green's Function. It is known from Green's theorem that if ϕ is a solution of the equation $(\nabla^2+k^2)\phi=0$, continuous along with its derivatives inside a closed curve C, and if Γ is another solution, continuous except near the point P where $\Gamma\sim-\log\rho$, ρ being the distance from P, then

$$\phi(P)=\frac{1}{2\pi}\int_C\left(\Gamma\frac{\partial\phi}{\partial n}-\phi\frac{\partial\Gamma}{\partial n}\right)ds, \quad . \quad . \quad . \quad . \quad . \quad . \qquad (31)$$

n being the outward drawn normal to C.

If, in vector notation, we denote by $\boldsymbol{r}$, $\boldsymbol{r}'$, the position of a general point in space, and the particular point P, respectively, then such a function satisfying the above conditions is

$$\Gamma(k|\boldsymbol{r}-\boldsymbol{r}'|)=\Gamma(x-x',\ y-y')=\frac{\pi i}{2}H_0^{(1)}(k|\boldsymbol{r}-\boldsymbol{r}'|), \quad . \quad . \qquad (32)$$

where $H_0^{(1)}(z)$ is the Bessel function of the third kind as defined by Nielson (Watson 1944). We note that the asymptotic form of $H_0^{(1)}(z)$ is, for large z,

$$H_0^{(1)}(z)\sim\left(\frac{2}{\pi z}\right)^{\frac{1}{2}}\exp\,[i(z-\pi/4)]. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (33)$$

We now take the contour of integration in (31) to be the curve s_1, s_2, s_3 shown in fig. 2. By virtue of the known asymptotic behaviour of ϕ_a

and Γ, the contributions from s_1 and s_3 vanish in the limit, and since $\partial\phi_a/\partial n=0$ along the walls of the channel, we obtain, finally, the relation

$$\phi_a(r)=\int_{-\infty}^{0}\mathrm{H}(x')\frac{\partial}{\partial y'}\Big[\Gamma(|\mathbf{r}-\mathbf{r}'|)_{y'=b}-\Gamma(|\mathbf{r}-\mathbf{r}'|)_{y'=-b}\Big]dx', \quad . \quad (34)$$

and the boundary condition that the normal derivative of ϕ_a should vanish at the walls of the channel gives

$$\int_{-\infty}^{0}\mathrm{H}(x')\frac{\partial^2}{\partial y\partial y'}\Big[\Gamma(|\mathbf{r}-\mathbf{r}'|)_{y'=b}-\Gamma(|\mathbf{r}-\mathbf{r}'|)_{y'=-b}\Big]_{y=b}dx'=0, \quad . \quad (35)$$

if $x<0$.

Equation (35) is the integral equation for $\mathrm{H}(x)$. For convenience it is written

$$\int_{-\infty}^{0}\mathrm{H}(x')\mathrm{K}(x-x')\,dx'=0, \quad x<0, \quad . \quad . \quad . \quad . \quad (36)$$

where

$$\mathrm{K}(x)=-\frac{1}{2\pi}\frac{\partial^2}{\partial y\partial y'}\Big[\Gamma(x, y-y')_{y'=b}-\Gamma(x, y-y')_{y'=-b}\Big]_{y=b}. \quad . \quad (37)$$

In deriving the above equation, no restriction to dominant mode propagation was made, and the result holds for an incident wave of arbitrary frequency. It is in determining the precise nature of the solution that the asymptotic behaviour of $\mathrm{H}(x)$ is required. In the present case the solution is subject to the condition that, for large x,

$$\mathrm{H}(x)\sim \mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx]. \quad . \quad . \quad . \quad . \quad (38)$$

which is sufficient to determine the function uniquely.

Analogous results were obtained by Levine and Schwinger (1948) for the propagation of sound waves in a circular pipe. It was shown by these authors that integral equations of the above type could be solved by the use of Fourier transforms and their analysis is here adapted to the two-dimensional channel.

SOLUTION OF THE INTEGRAL EQUATION.

In solving (36), the equation is first written in the more general form

$$\left.\begin{aligned}\int_{-\infty}^{0}\mathrm{H}(x')\mathrm{K}(x-x')\,dx'&=\mathrm{E}(x), \quad x\geqslant 0\\ &=0, \qquad\quad x<0,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (39)$$

so that, multiplying each side by $\exp[-i\zeta x]$, and integrating over all x, we obtain

$$\int_{-\infty}^{0}\mathrm{H}(x')\,dx'\int_{-\infty}^{\infty}\mathrm{K}(x-x')\exp[-i\zeta x]\,dx=\int_{0}^{\infty}\mathrm{E}(x)\exp[-i\zeta x]\,dx,$$

or

$$\int_{-\infty}^{0}\mathrm{H}(x')\exp[-i\zeta x']\,dx'\int_{-\infty}^{\infty}\mathrm{K}(z)\exp[-i\zeta z]\,dz=\int_{0}^{\infty}\mathrm{E}(x)\exp[-i\zeta x]\,dx.$$

i. e.

$$\mathrm{H}(\zeta)\mathrm{K}(\zeta)=\mathrm{E}(\zeta), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

$\mathrm{E}(\zeta)$ being the transform of the right-hand side of (39).

In the application of equation (40) it is necessary that the functions occurring should have a common region of regularity. We show that this is true for $|\eta|<\epsilon$ where $\zeta=\xi+i\eta$ and $\mathrm{I}mk=\epsilon$.

From the asymptotic form of $\mathrm{H}(x)$ we know that $\mathrm{H}(\zeta)$ is regular for $\eta>\epsilon$ but has simple poles at $\zeta=\pm k$. However, by inserting the factor (ζ^2-k^2), we may take $(\zeta^2-k^2)\mathrm{H}(\zeta)$ to be regular for $\eta\geqslant-\epsilon$.

Moreover, by inserting the asymptotic form of Γ in equation (37) we obtain

$$\mathrm{K}(x)\sim\left(\frac{\pi}{2k}\right)^{\frac{1}{2}}\frac{\exp\left[i(k|x|+\pi)\right]}{|x|^{\frac{3}{2}}} \qquad (41)$$

so that $\mathrm{K}(\zeta)$ is regular for $|\eta|<\epsilon$.

The explicit relation for $\mathrm{K}(\zeta)$ can be obtained from that of $\Gamma(\zeta, y-y')$, which is shown in Appendix I. to be

$$\Gamma(\zeta, y-y')=\frac{i\pi\exp\left[i(k^2-\zeta^2)^{\frac{1}{2}}|y-y'|\right]}{(k^2-\zeta^2)^{\frac{1}{2}}}, \qquad (42)$$

where the argument of $(k^2-\zeta^2)^{\frac{1}{2}}$ is taken to be $\pi/2$ as $|\zeta|\rightarrow\infty$ on the real axis.

It follows, from equation (37), that

$$\mathrm{K}(\zeta)=b(k^2-\zeta^2)\mathrm{L}(\zeta), \qquad (43)$$

where

$$\mathrm{L}(\zeta)=-\frac{\left[1-\exp\left\{2bi(k^2-\zeta^2)^{\frac{1}{2}}\right\}\right]}{2bi(k^2-\zeta^2)^{\frac{1}{2}}} \qquad (44)$$

and $\mathrm{L}(\zeta)$ is regular for $|\eta|<\epsilon$. Moreover $\mathrm{L}(\zeta)$ is non-zero inside the strip; its zeros occur at $\zeta=\pm i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}$ and these points lie outside the strip.

Thus the left-hand side of (40) is composed of two factors, $\mathrm{L}(\zeta)$ and $b(k^2-\zeta^2)\mathrm{H}(\zeta)$, for which $|\eta|<\epsilon$ is a common region of regularity.

Finally we note, from (34) and (37) and the definition of $\mathrm{E}(x)$ contained in (39), that $\mathrm{E}(x)$ represents the y-derivative of the potential at $y=b$. Its asymptotic behaviour can therefore be derived from that of ϕ_a outside the channel. With the help of (4) we obtain, for large x,

$$\mathrm{E}(x)\sim-\tfrac{1}{2}bf(0)\frac{\exp\left[ikx\right]}{x^{3/2}}, \qquad (45)$$

so that $\mathrm{E}(\zeta)$ is regular for $\eta\leqslant\epsilon$.

The next step in the solution of (39) is to identify $\mathrm{L}(\zeta)$ as the quotient of two functions, $\mathrm{L}_+(\zeta)$ and $\mathrm{L}_-(\zeta)$, which are regular and non-zero for $\eta>-\epsilon$, $\eta<\epsilon$, respectively. Equation (40) will then become

$$b(k^2-\zeta^2)\mathrm{H}(\zeta)\mathrm{L}_+(\zeta)=\mathrm{E}(\zeta)\mathrm{L}_-(\zeta), \qquad (46)$$

the left-hand side being regular for $\eta>-\epsilon$ and the right-hand side for

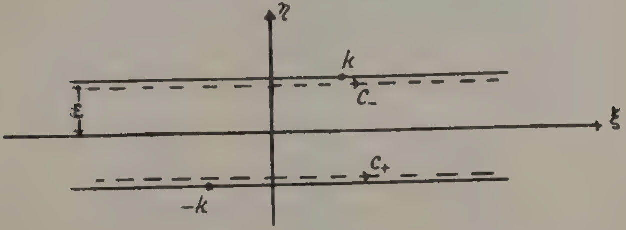
$\eta<\epsilon$. By analytic continuation they define, together, a function regular in the whole finite domain. It will be shown in Appendix II. that the integral function so defined is a constant so that a knowledge of $L_+(\zeta)$ implies a knowledge of $H(\zeta)$ apart from a constant factor. Reference to equations (26) and (29) shows that the expressions for the reflection coefficient R, and the function $G(\theta)$ describing the transmitted energy, are independent of such a factor multiplying $H(\zeta)$ so that its explicit evaluation is unnecessary.

The functions $L_+(\zeta)$ and $L_-(\zeta)$ are obtained by applying Cauchy's formula to $\log L(\zeta)$. Since $L(\zeta)$ is regular and non-zero in the region $|\eta|<\epsilon$, we can write, for $|\eta|<\epsilon$,

$$\log L(\zeta)=\frac{1}{2\pi i}\oint\frac{\log L(t)\,dt}{t-\zeta}, \quad \ldots \ldots \quad (47)$$

the integration being taken round an elongated rectangle within the strip and on the branch for which $\log 1=0$ (fig. 3).

Fig. 3.



From (44) we find that $\log L(t)\sim\log|t|$ as $|t|\rightarrow\infty$ within the strip. Consequently the contributions from the vertical sides of the rectangle tend to zero as the contour extends to infinity and we have the relation

$$\log L(\zeta)=\frac{1}{2\pi i}\int_{C_+}\frac{\log L(t)\,dt}{t-\zeta}-\frac{1}{2\pi i}\int_{C_-}\frac{\log L(t)}{t-\zeta} \quad \ldots \quad (46)$$

$$=\log L_+(\zeta)-\log L_-(\zeta). \quad \ldots \ldots \quad (47)$$

The contours C_+ and C_- are shown in fig. 3, and, to ensure the individual convergence of L_+ and L_- we take

$$\int_{C_+}\frac{\log L(t)\,dt}{t-\zeta}=\lim_{T\rightarrow\infty}\int_{-T-i\delta}^{T-i\delta}\frac{\log L(t)\,dt}{t-\zeta}, \quad \ldots \quad (48)$$

where $0<\delta<\epsilon$.

The second integral is defined in a similar manner.

The two functions $L_+(\zeta)$ and $L_-(\zeta)$ so defined are regular and non-zero in the regions $\eta>-\delta$ and $\eta<\delta$ respectively. In the limit when the two contours approach the axis, the two functions L_+ and L_- are represented by the common expression

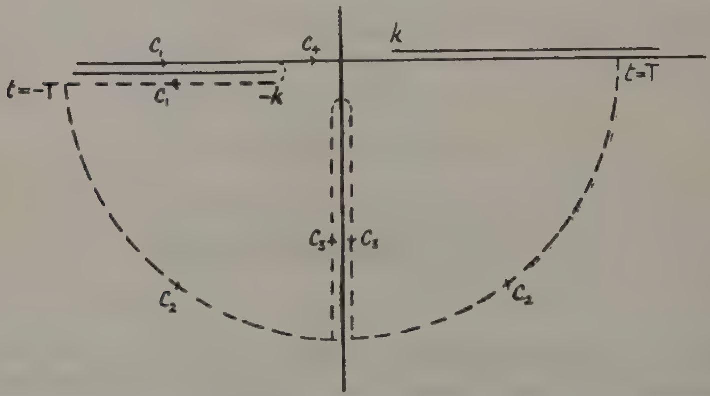
$$\exp\left[\frac{1}{2\pi i}\lim_{T\to\infty}\int_{-T}^{T}\frac{\log L(t)\,dt}{t-\zeta}\right]$$
$$=\exp\left[\frac{1}{2\pi i}\lim_{T\to\infty}\int_{-T}^{T}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}\right] \quad . \quad . \quad (49)$$

and

$$L_+(-\zeta)=\frac{1}{L_-(\zeta)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

For real k (*i. e.* as $\epsilon\to 0$), the contours C_+, C_- should pass the branch point $t=k$ on the lower side of the real axis and $t=-k$ on the upper side. The two contours then lie wholly in the respective regions of regularity of the two functions L_+, L_- and the phase of $(k^2-t^2)^{\frac{1}{2}}$ changes from 0 to $\pi/2$ for $|t|<k$, $|t|>k$ respectively.

Fig. 4.



The above expression for L_+ could now be used to deduce the function $H(\zeta)$ apart from a constant factor, and hence the asymptotic behaviour of the reflected and transmitted waves explicitly. It is possible, however, to express $L_+(\zeta)$ in a somewhat more significant form which is most easily obtained by a deformation of the contour C_+. The discussion of convergence is also simplified if we proceed from the derivative of $\log L_+(\zeta)$ which is

$$\frac{d}{d\zeta}\log L_+(\zeta)=\frac{1}{2\pi i}\int_{C_+}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{(t-\zeta)^2}. \quad . \quad . \quad (51)$$

Remembering that ζ lies in the upper half plane, we deform the contour C_+ into the contour $C_1\,C_2\,C_3$ shown in fig. 4. The section C_1 is the upper and lower sides of the branch line extending from $-\infty$ to $-k$; C_2 consists of two arcs of a circle, radius T, in the lower half-plane, while C_3 excludes the singularities of the integrand in (51) at the zeros of $\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}$ on the imaginary axis.

Along the upper and lower sides of the branch cut, the argument of $(k^2-t^2)^{\frac{1}{2}}$ is $+\pi/2$ and $-\pi/2$ respectively, so that the contribution to the right-hand side of equation (51) from C_1 is

$$\frac{1}{2\pi i}\int_{C_1}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{(t-\zeta)^2}$$
$$=\frac{1}{2\pi i}\int_k^{T}\log\left[\frac{1-\exp[-2b(t^2-k^2)^{\frac{1}{2}}]}{-(1-\exp[-2b(t^2-k^2)^{\frac{1}{2}}])}\right]\frac{dt}{(t+\zeta)^2}=-\frac{b}{\pi i}\int_k^{T}\frac{(t^2-k^2)^{\frac{1}{2}}\,dt}{(t+\zeta)^2}$$
$$=-\frac{b}{\pi i}\left[-\frac{(T^2-k^2)^{\frac{1}{2}}}{T+\zeta}+\cosh^{-1}T/k-\zeta\int_0^{\cosh^{-1}T/k}\frac{du}{k\cosh u+\zeta}\right]\quad . \quad . \quad . \quad (52)$$

so that

$$\lim_{T\to\infty}\left[\frac{1}{2\pi i}\int_{C_1}+\frac{b}{\pi i}\cosh^{-1}\frac{T}{k}\right]=\frac{b}{\pi i}\left[1+\frac{2\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]. \quad . \quad (53)$$

Consider now the integral round C_2 in the quadrant for which $-\pi<\arg t<-\pi/2$. In this range $-\pi/2<\arg(k^2-t^2)^{\frac{1}{2}}<0$, and the asymptotic value of the integrand in (51) is $-(\log 2bt)/t^2$ so that, in the limit $T\to\infty$, the integral along C_2 is zero in this quadrant.

In the range $-\pi/2<t<0$, however, we have $0<\arg(k^2-t^2)^{\frac{1}{2}}<\pi/2$. On this section of C_2 the asymptotic value of the integrand is $(\log\exp[-2bt])/t^2=-2b/t$ and the contribution is therefore

$$-\frac{2b}{2\pi i}\int\frac{dt}{t}=-\frac{b}{\pi i}\int_{-\pi/2}^{0}d\theta=-\frac{b}{2}. \quad . \quad . \quad . \quad . \quad . \quad (54)$$

Finally, there is the contribution from C_3 due to the phase difference of the integrand on the two sides of the imaginary axis. Since the zeros of $[1-\exp\{2bi(k^2-t^2)^{\frac{1}{2}}\}]$ occur at $t=-i(n^2\pi^2/b^2-k^2)^{\frac{1}{2}}$, it follows that the integrand in (51) has the same phase, along that part of the negative real axis for which $|Im t|<T$, as

$$\prod_{n=1}^{N}\left[t+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}\right],$$

where N is the largest integer not exceeding Tb/π. It follows that the integral along the two sides of the negative real axis is

$$\frac{1}{2\pi i}\int_{C_3}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{(t-\zeta)^2}$$
$$=\frac{1}{2\pi i}\int_{C_3}\log\prod_{n=1}^{N}\left\{t+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}\right\}\frac{dt}{(t-\zeta)^2}$$
$$=-\sum_{n=1}^{N}\int_{-i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}^{-iT}\frac{dt}{(t-\zeta)^2}=-\sum_{n=1}^{N}\left[\frac{1}{\zeta+iT}-\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}\right]$$
$$. \quad . \quad . \quad (55)$$

If the contributions from the various sections of contour, given by (53), (54), (55), are now added, we obtain

$$\frac{d}{d\zeta}\log \mathrm{L}_{+}(\zeta)=\frac{b}{\pi i}\left[1+\frac{2\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]-\frac{b}{2}$$
$$-\lim_{\mathrm{T}\to\infty}\left[\sum_{n=1}^{\mathrm{N}}\left\{\frac{1}{\zeta+i\mathrm{T}}-\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}+\frac{b}{\pi i}\cosh^{-1}\frac{\mathrm{T}}{k}\right\}\right]. \quad . \quad . \quad . \quad (56)$$

Now, since, for large z and N,

$$\cosh^{-1}z\sim\log 2z$$

$$\sum_{n=1}^{\mathrm{N}}\frac{1}{n}\sim\log \mathrm{N}+\gamma,$$

where γ is Euler's constant, it follows that (56) can be written

$$\frac{d}{d\zeta}\log \mathrm{L}_{+}(\zeta)=\frac{b}{\pi i}\left[1+\frac{2\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]-\frac{b}{2}$$
$$+\lim_{\mathrm{T}\to\infty}\left[\sum_{n=1}^{\mathrm{N}}\left\{\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{1}{\zeta+i\mathrm{T}}-\frac{b}{n\pi i}\right\}\right.$$
$$\left.+\frac{b\gamma}{\pi i}+\frac{b}{\pi i}\log \mathrm{N}-\frac{b}{\pi i}\log\frac{2\mathrm{T}}{k}\right]=\frac{2\zeta b}{\pi i(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}-\frac{b}{2}$$
$$+\sum_{n=1}^{\infty}\left\{\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{b}{n\pi i}\right\}+\frac{b\gamma}{\pi i}+\frac{b}{\pi i}\log\frac{bk}{2\pi}, \quad . \quad (57)$$

since $\mathrm{N}\sim\mathrm{T}b/\pi$.

A simple integration with respect to ζ then gives

$$\mathrm{L}_{+}(\zeta)=\kappa\exp\left[-\frac{b\zeta}{2}+\frac{i\zeta b}{\pi}\log\frac{2\pi}{bk}-\frac{i\zeta b\gamma}{\pi}+\frac{i\zeta b}{\pi}+\frac{2ib}{\pi}(k^2-\zeta^2)^{\frac{1}{2}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right.$$
$$\left.+\sum_{n=1}^{\infty}\left\{\log\left[\frac{\zeta b}{n\pi}+i\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}\right]+\frac{ib\zeta}{n\pi}\right\}\right]$$
$$=\kappa'\prod_{n=1}^{\infty}\left[\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}-\frac{i\zeta b}{n\pi}\right]\exp\left(\frac{i\zeta b}{n\pi}\right)\exp\left[\frac{i\zeta b}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-\frac{b\zeta}{2}\right.$$
$$\left.+\frac{2ib}{\pi}(k^2-\zeta^2)^{\frac{1}{2}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

Since

$$\mathrm{L}_{+}(-\zeta)=\frac{1}{\mathrm{L}_{-}(\zeta)},$$

equation (58) gives

$$\mathrm{L}_{+}(\zeta)/\mathrm{L}_{-}(\zeta)=\mathrm{L}(\zeta)=(\kappa')^2\prod_{n=1}^{\infty}\left[1-\frac{(k^2-\zeta^2)b^2}{n^2\pi^2}\right]\exp\left[ib(k^2-\zeta^2)^{\frac{1}{2}}\right]$$
$$=\frac{(\kappa')^2 i}{2b(k^2-\zeta^2)^{\frac{1}{2}}}\left(1-\exp\left[2bi(k^2-\zeta^2)^{\frac{1}{2}}\right]\right), \quad . \quad . \quad . \quad (59)$$

and comparison of equation (59) with (44) shows that the integration constant $\kappa'=1$.

Moreover

$$\prod_{n=1}^{\infty}\left\{\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}-\frac{i\zeta b}{n\pi}\right\}\exp\left[\frac{i\zeta b}{n\pi}\right]$$

$$=\exp\left[\sum_{n=1}^{\infty}\log\left\{\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}-\frac{i\zeta b}{n\pi}\right\}+\frac{i\zeta b}{n\pi}\right]$$

$$=\exp\left[\tfrac{1}{2}\sum_{n=1}^{\infty}\log\left\{1-\frac{(k^2-\zeta^2)b^2}{n^2\pi^2}\right\}-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right]$$

$$=\prod_{n=1}^{\infty}\left\{1-\frac{(k^2-\zeta^2)b^2}{n^2\pi^2}\right\}^{\frac{1}{2}}\exp\left[-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right]$$

$$=\left[\frac{\sin b(k^2-\zeta^2)^{\frac{1}{2}}}{b(k^2-\zeta^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}}\exp\left[-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right], \quad . \; . \; . \quad (60)$$

so that

$$L_+(\zeta)=\left[\frac{\sin b(k^2-\zeta^2)^{\frac{1}{2}}}{b(k^2-\zeta^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}}\exp\left[-\frac{b\zeta}{2}+\frac{i\zeta b}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)\right.$$
$$\left.+\frac{2ib}{\pi}(k^2-\zeta^2)^{\frac{1}{2}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right], \quad . \; . \; . \quad (61)$$

and

$$L_+(k\cos\theta)=\left[\frac{\sin(bk\sin\theta)}{bk\sin\theta}\right]^{\frac{1}{2}}\exp\left[-\frac{bk\cos\theta}{2}+\frac{ibk\cos\theta}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)\right.$$
$$\left.+\frac{ibk\theta\sin\theta}{\pi}-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{k\cos\theta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{bk\cos\theta}{n\pi}\right\}\right]. \quad . \; . \quad (62)$$

$$L_+(k)=\exp\left[-\frac{bk}{2}+\frac{ibk}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-i\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{bk}{n\pi}-\frac{bk}{n\pi}\right\}\right]. \quad . \; . \quad (63)$$

It is now a simple matter to obtain expressions describing the reflected and transmitted waves. The reflection coefficient is, from equation (26),

$$R=\operatorname*{Res}_{\zeta=-k}H(\zeta)\Big/\operatorname*{Res}_{\zeta=k}H(\zeta), \quad . \; . \; . \; . \; . \; . \; . \quad (64)$$

and since, (see Appendix)

$$(k^2-\zeta^2)\mathrm{H}(\zeta)\mathrm{L}_+(\zeta)=\mathrm{D}, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (65)$$

where D is some constant, it follows that

$$\underset{\zeta=\pm k}{\mathrm{Res}}=\pm\frac{\mathrm{D}}{2k\mathrm{L}_+(\pm k)}. \quad . \; . \; . \; . \; . \; . \; . \quad (66)$$

Equation (64) can now be written

$$\mathrm{R}=-|\mathrm{R}|\exp[2ikl]=-\mathrm{L}_+(k)/\mathrm{L}_+(-k)=-\{\mathrm{L}_+(k)\}^2, \quad . \quad (67)$$

where l is the end correction.

Using the expression for $\mathrm{L}_+(k)$ given in equation (63), we find that

$$|\mathrm{R}|=\exp[-bk], \quad . \; . \; . \; . \; . \; . \; . \; . \quad (68)$$

$$\frac{l}{b}=\frac{1}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-\frac{1}{bk}\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{bk}{n\pi}-\frac{bk}{n\pi}\right\}. \quad . \; . \; . \quad (69)$$

Finally we derive the expression for $\mathrm{G}(\theta)$. Equation (65) shows that

$$\mathrm{H}(k\cos\theta)=\frac{\mathrm{D}}{(k\sin\theta)^2\mathrm{L}_+(k\cos\theta)}. \quad . \; . \; . \; . \; . \; . \quad (70)$$

Substituting (66) and (70) in (29) we obtain

$$\mathrm{G}(\theta)=\frac{2\sin^2(bk\sin\theta)}{bk\sin^2\theta}\,\frac{|\mathrm{L}_+(k)|^2}{(1-|\mathrm{R}|)^2|\mathrm{L}_+(k\cos\theta)|^2} \quad . \; . \; . \quad (71)$$

$$=\frac{\exp[bk\cos\theta]\sin(bk\sin\theta)}{\sin\theta\sinh bk}, \quad . \; . \; . \; . \; . \; . \quad (72)$$

by virtue of relations (62), (63) and (68).

A simple check on this last relation is obtained by noting that from its definition the function $\mathrm{G}(\theta)$ must satisfy

$$\int_{-\pi}^{\pi}\mathrm{G}(\theta)\,d\theta=2\pi. \quad . \; . \; . \; . \; . \; . \; . \; . \quad (73)$$

To verify this we consider the integral

$$2\oint\frac{\exp[bkz]}{z^2-1}\,dz, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (74)$$

taken round the unit circle in the z-plane, with indentations at the poles $z=\pm 1$.

The integrand is free from singularities inside the contour and the integral is therefore zero. If we substitute $z=\exp[i\theta]$, then this is equivalent to

$$\mathrm{P}\int_{-\pi}^{\pi}\exp[bk\exp(i\theta)]\frac{d\theta}{\sin\theta}-\pi i\left[\underset{z=1}{\mathrm{Res}}+\underset{z=-1}{\mathrm{Res}}\right]=0 \quad . \; . \quad (75)$$

or

$$\int_{-\pi}^{\pi} \exp\,[bk\cos\theta]\,\frac{\sin\,(bk\sin\theta)}{\sin\theta}\,d\theta=\pi(\exp\,[bk]+\exp\,[-bk]) \tag{76}$$

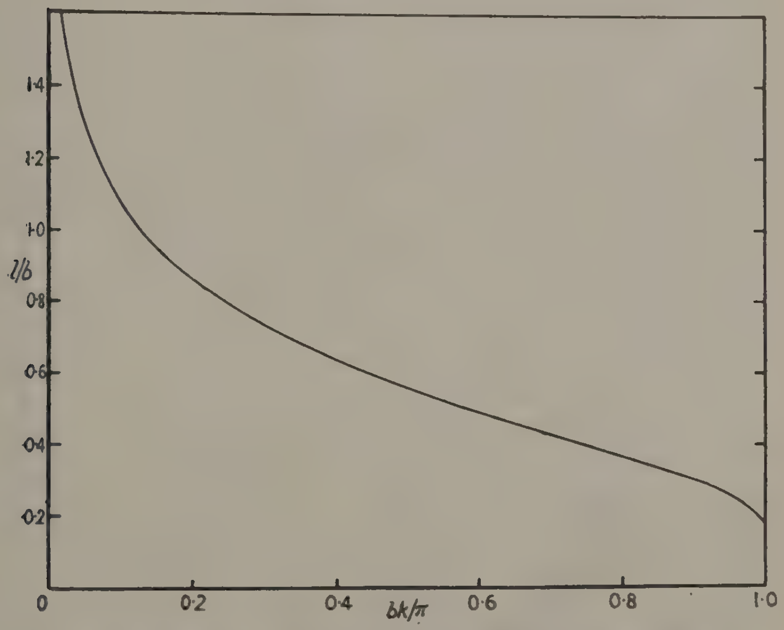
so that

$$\int_{-\pi}^{\pi} \mathrm{G}(\theta)\,d\theta=\int_{-\pi}^{\pi} \exp\,[bk\cos\theta]\,\frac{\sin\,(bk\sin\theta)}{\sin\theta\,\sinh bk}\,d\theta=2\pi. \quad . \quad . \tag{77}$$

DISCUSSION OF RESULTS.

One of the most striking features is the extremely simple relation for the amplitude of the reflected wave (equation (68)). The corresponding result for the circular pipe is more complex, but, for large values of the wavelength, the amplitude is approximately proportional to $\exp\,[-k^2c^2/2]$ where c is the radius of the pipe (Levine and Schwinger 1948). The decrease amplitude with the wavelength, when the latter is large, is therefore smaller than in the two-dimensional case.

Fig. 5.



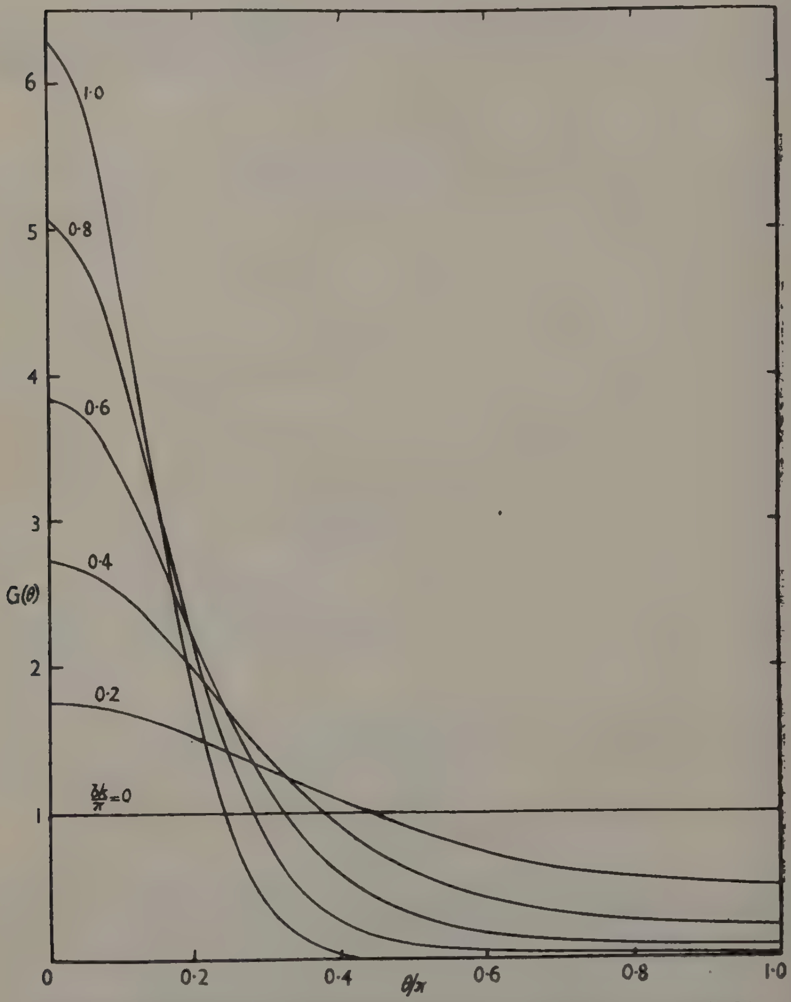
The variation in end-correction as a function of $\dfrac{bk}{\pi}\left(=\dfrac{2b}{\lambda}\right)$.

The end correction is illustrated graphically in fig. 5. Equation (69) shows that it increases indefinitely as the wavelength becomes infinitely great, as opposed to the circular pipe for which it tends to a finite limit. The same conclusion was reached by Rayleigh (1904) using approximate methods. The physical significance of the end-correction is consequently somewhat remote and it is more constructive to consider the actual correction to the phase in the returning wave which is given by

$$2kl=\frac{2bk}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-2\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{bk}{n\pi}-\frac{bk}{n\pi}\right\} \quad . \quad . \quad (78)$$

and tends to zero for large values of the wavelength.

Fig. 6.



The radiated energy as a function of θ/π for various values of bk/π.

In fig. 6, the variation in G(θ) is shown for various values of $bk/\pi(=2b/\lambda)$. When the wavelength is comparable with the width of the channel, most of the energy is transmitted in a forward direction. The function decreases monotonically with θ except for values of bk/π near unity when there is a minimum in the region of $\theta=\pi/2$. If $bk=\pi$, for example, $G(\pi/2)=0$ and no energy is propagated in this direction. Beyond this value $G(\theta)$ increases slowly to 0·0118.

APPENDIX I.

To derive the Fourier transform of

$$\Gamma(x, y)=\frac{\pi i}{2} H_0^{(1)}[k(x^2+y^2)^{\frac{1}{2}}] \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{A. }1)$$

we begin from the known relation (Watson 1944)

$$H_0^{(1)}(z)=\frac{1}{\pi i}\int_{-\infty}^{\infty} \exp\,[iz\cosh t]\,dt \text{ if } 0\leqslant\arg z\leqslant\pi, \quad . \quad . \quad (\text{A. }2)$$

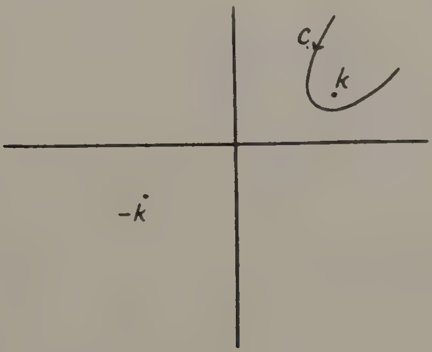
so that

$$\Gamma(x, y)=\tfrac{1}{2}\int_{-\infty}^{\infty} \exp\,[ik(x^2+y^2)^{\frac{1}{2}}\cosh t]\,dt \text{ if } \mathrm{I}mk=\epsilon>0$$

$$=\tfrac{1}{2}\int_{-\infty-i\theta}^{\infty-i\theta} \exp\,[ikx\cosh u-k|y|\sinh u]\,du, \quad . \quad (\text{A. }3)$$

where $\tan\theta=|y|/z$.

Fig. 7.



Substituting $\zeta=k\cosh u$ this becomes

$$\Gamma(x, y)=-\frac{1}{2i}\int\frac{\exp[ix\zeta+i|y|(k^2-\zeta^2)^{\frac{1}{2}}]\,d\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}, \quad . \quad . \quad (\text{A. }4)$$

where the contour of integration begins and ends at infinity and passes between the two branch points $\zeta=\pm k$ (see fig. 7). By a deformation of the contour, this can be written, if $|\eta|<\epsilon$

$$\Gamma(x, y)=-\frac{1}{2i}\int_{-\infty+i\eta}^{\infty+i\eta}\frac{\exp\,[ix\zeta+i|y|(k^2-\zeta^2)^{\frac{1}{2}}]\,d\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}. \quad . \quad (\text{A. }5)$$

It follows that the transform of $\Gamma(x, y)$ is $\Gamma(\zeta, y)$ where

$$\Gamma(\zeta, y)=\frac{i\pi \exp\left[i(k^2-\zeta^2)^{\frac{1}{2}}\,|y|\right]}{(k^2-\zeta^2)^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad (\text{A. }6)$$

and is regular for $|Im\zeta|<\epsilon$.

APPENDIX II.

During the course of the analysis, use was made of the fact that the integral function defined by the two sides of the equation

$$b(k^2-\zeta^2)^{\frac{1}{2}}\mathrm{H}(\zeta)\mathrm{L}_+(\zeta)=\mathrm{E}(\zeta)\mathrm{L}_-(\zeta)$$

was, in fact, a constant.

To prove this we examine the asymptotic values of the two sides, in their respective regions of regularity, as $|\zeta|\rightarrow\infty$.

For the asymptotic behaviour of $\mathrm{L}_+(\zeta)$ we refer to equation (51),

$$\begin{aligned}\mathrm{L}_+(\zeta)&=\exp\left[\frac{1}{2\pi i}\lim_{\mathrm{T}\rightarrow\infty}\int_{-\mathrm{T}}^{\mathrm{T}}\log\left\{\frac{i\{1-\exp\left[2bi(k^2-t^2)^{\frac{1}{2}}\right]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}\right]\\&=\exp\left[\frac{1}{2\pi i}\int_{-k}^{k}\log\left\{\frac{i\{1-\exp\left[2bi(k^2-t^2)^{\frac{1}{2}}\right]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}\right.\\&\quad\left.+\frac{\zeta}{\pi i}\int_{k}^{\infty}\log\left\{\frac{1-\exp\left[-2b(t^2-k^2)^{\frac{1}{2}}\right]}{2b(t^2-k^2)^{\frac{1}{2}}}\right\}\frac{dt}{t^2-\zeta^2}\right].\quad . \quad . \quad . \quad (\text{A. }7)\end{aligned}$$

Now

$$\frac{1}{2\pi i}\int_{-k}^{k}\log\left\{\frac{i\{1-\exp\left[2bi(k^2-t^{2\frac{1}{2}})\right]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}=-0\left(\frac{1}{\zeta}\right) \quad . \quad (\text{A. }8)$$

and

$$\begin{aligned}\int_{k}^{\infty}\log\left\{\frac{1-\exp\left[-2b(t^2-k^2)^{\frac{1}{2}}\right]}{2b(t^2-k^2)^{\frac{1}{2}}}\right\}\frac{dt}{t^2-\zeta^2}&=\int_{k}^{\infty}\log\left\{\frac{t\{1-\exp\left[-2b(t^2-k^2)^{\frac{1}{2}}\right]\}}{2b(t^2-k^2)^{\frac{1}{2}}}\right\}\\&\quad\times\frac{dt}{t^2-\zeta^2}+\int_{k}^{\infty}\log\left(\frac{1}{t}\right)\frac{dt}{t^2-\zeta^2}\\&=-0\left(\frac{1}{\zeta^2}\right)+\frac{i}{\zeta}\int_{-k/i\zeta}^{\infty}\log\left(\frac{1}{-i\zeta v}\right)\frac{dv}{1+v^2}\\&\quad . \quad . \quad . \quad (\text{A. }9)\\&\sim\frac{\pi i}{2\zeta}\log\left(\frac{1}{-i\zeta}\right)+0\left(\frac{1}{\zeta}\right). \quad . \quad (\text{A. }10)\end{aligned}$$

Substituting in (A. 7) we obtain

$$\mathrm{L}_+(\zeta)\sim\frac{1}{(-i\zeta)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{A. }11)$$

and

$$\mathrm{L}_-(\zeta)=\frac{1}{\mathrm{L}_+(-\zeta)}\sim(i\zeta)^{\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{A. }12)$$

For the asymptotic behaviour of

$$H(\zeta)=\int_{-\infty}^{0} H(x)\exp[-i\zeta x]\,dx \quad Im\zeta>0, \quad . \quad . \quad (A.\ 13)$$

and

$$E(\zeta)=\int_{0}^{\infty} E(x)\exp[-i\zeta x]\,dx \quad Im\zeta<0, \quad . \quad . \quad . \quad (A.\ 14)$$

we note that the significant contributions to the integrals, when $|\zeta|$ is large, will arise from the values of $H(x)$, $E(x)$ for small negative and positive x respectively. Now the difference in potential across the plane $y=b$ must be a continuous function of x which vanishes for $x>0$. It follows that, as $x\rightarrow -0$

$$H(x)\sim x^{\alpha}, \quad \alpha>0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A.\ 15)$$

or

$$H(\zeta)\sim(-i\zeta)^{-\alpha-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A.\ 16)$$

Again the y-derivative of the potential across the plane $y=b$ must have an integrable singularity at $x=0$, and since this is just $E(x)$, $x>0$, we have, as $x\rightarrow +0$

$$E(x)\sim x^{-\beta}, \quad \beta<1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A.\ 17)$$

or

$$E(\zeta)\sim(i\zeta)^{\beta-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A.\ 18)$$

Using (A. 11), (A. 12), (A. 13) and (A. 14) we see that the behaviour of the integral function is, in the upper half plane

$$b(k^2-\zeta^2)H(\zeta)L_{+}(\zeta)\sim(-i\zeta)^{\frac{1}{2}-\alpha},$$

while in the lower half plane

$$E(\zeta)L_{-}(\zeta)\sim(i\zeta)^{\beta-\frac{1}{2}}.$$

Because of the limits on α and β, the integral function cannot tend to infinity faster than $\zeta^{\frac{1}{2}}$ and must therefore be a constant.

REFERENCES.

CARSLAW, 1898, *Proc. Lond. Math. Soc.*, **30**, 121.
LEVINE and SCHWINGER, 1948, *Phys. Rev.*, **73**, 383.
RAYLEIGH, 1904, *Phil. Mag.*, **8**, 481.
RAYLEIGH, 1940, *Theory of Sound* (London : Macmillan & Co.), **2**, chap. XVI.
SOMMERFELD, 1894, *Math. Ann.*, **47**, 368.
Titchmarsh, 1937, *Theory of Fourier Integrals* (Oxford : University Press).
WATSON, 1944, *Theory of Bessel Functions* (Cambridge : University Press).

III. *Betatron-Starting in Electron Synchrotrons.*

By J. J. Wilkins,
Atomic Energy Research Establishment, Harwell*.

[Received September 19, 1949.]

Abstract.

The magnetic design of betatron-starting flux bars for electron synchrotrons is considered in relation to the principles governing efficient initial electron injection and efficient betatron-synchrotron transition. A brief discussion of betatron-synchrotron transition theory is included, in order to show how flux bar performance may be specified therefrom. A design procedure is derived for determining the essential flux bar dimensions, approximate corrections being made for skin effect in the laminated material of the bars.

To illustrate the principles discussed in the main paper, the results of some detailed performance calculations on the 375 MeV. Glasgow University synchrotron are given in an Appendix.

§1. Introduction.

In designing the magnet of a betatron-started electron synchrotron, economy in weight and space dictates the use of betatron flux bars of the minimum size consistent with satisfactory performance. Essentially this implies that the betatron acceleration phase of the machine is restricted to imparting just sufficient electron energy to ensure efficient transition to synchrotron acceleration. The betatron bars may then saturate, while synchrotron acceleration maintains the energy increase.

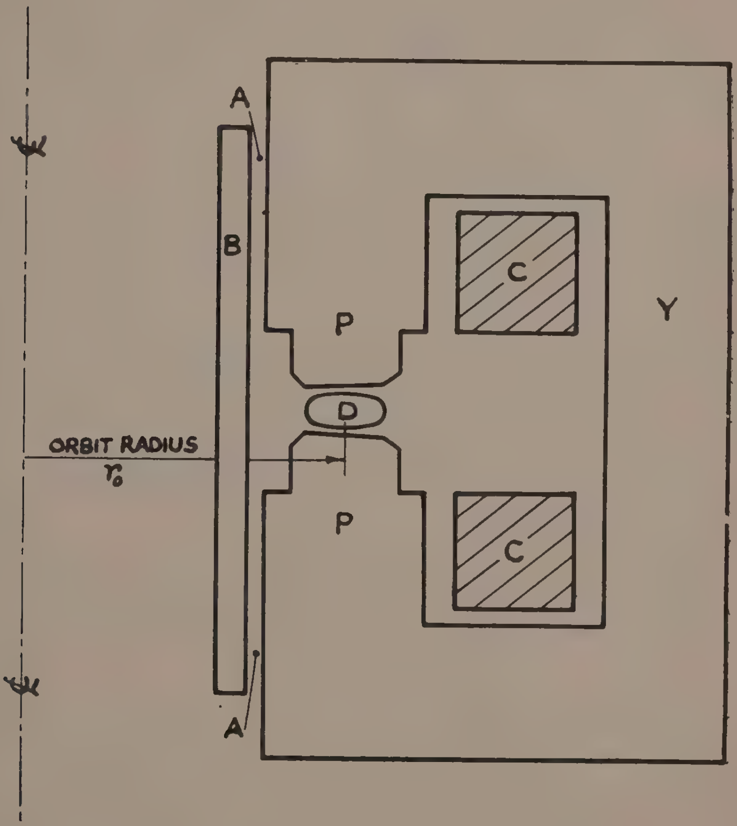
Two fundamental criteria govern satisfactory betatron bar performance. The first is that, to obtain efficient electron injection into the betatron phase, the bars must produce an equilibrium orbit located centrally in the vacuum donut. The second is that, during the betatron phase, the electron orbits must be maintained sufficiently central in the donut to permit efficient transition to synchrotron acceleration when the R.F. accelerating voltage is applied. As Goward (1949) has shown, this second criterion involves marked differences between machines of various energies.

For the sake of definiteness, attention in this paper is focused primarily on synchrotrons producing from about 100 to 400 MeV., and having a general magnet arrangement of the type illustrated in fig. 1. With the exception of §§3 and 4, however, the treatment throughout applies, without essential modification, to other energy ranges and magnet designs. In §2 it is shown how the first performance criterion mentioned above determines the dimensions of the air gap (or gaps) controlling the betatron flux. The second criterion, namely efficient betatron-synchrotron transition, effectively determines the minimum permissible betatron bar cross section. §§3 and 4 show how the simple transition theory given by Goward (1949) may be put into a convenient form for specifying betatron bar performance, treating the problem in full only for synchrotrons of the energy range mentioned above (100 to 400 MeV.). §§5 and 6

---

* Communicated by Denis Taylor.

Fig. 1.



General arrangement of a synchrotron magnet.
A. Air gaps controlling betatron flux.
B. Betatron flux bar.
C. Main excitation coils.
D. Vacuum donut.
P. Magnet poles.
Y. Magnet return path.

then give a method for determining an appropriate betatron bar cross section, employing the D.C. hysteresis loop of the bar laminations, and §7 shows how a moderate amount of "skin effect" in the laminations may be allowed for. No account has been attempted of the effects of cross fluxes in any part of the magnetic circuits, although in practice such fluxes are difficult to avoid entirely.

In Appendix I electrical methods of modifying the betatron-starting flux in a synchrotron are briefly reviewed. In Appendix II, some detailed performance characteristics of the 375 MeV. Glasgow University synchrotron are presented, illustrating various topics discussed in the main paper. A particular attempt is made to estimate overall electron trapping efficiencies more accurately than is possible by Goward's much simplified theory.

§2. BETATRON AIR GAP DESIGN.

(i) *Fundamental Equilibrium Orbit Equations.*

To establish an equilibrium orbit of radius r_0 in a betatron, the well known Wideröe flux condition must be satisfied :

$$\partial\phi_0/\partial t = 2\pi r_0^2 \partial \mathrm{H}_0/\partial t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

H_0 is the vertical field at radius r_0 in the annular magnet gap, and ϕ_0 is the total flux linking the orbit. Internal leakage fluxes of the synchrotron magnet, together with the flux in the annular gap at radii less than r_0, contribute an appreciable fraction of ϕ_0, and may be expressed as $(1-k_0)2\pi r_0^2\mathrm{H}_0$, where, for a given magnet, k_0 depends only on r_0. The remainder of ϕ_0 is the flux ϕ_b carried by the betatron bars. Thus ϕ_b and r_0 are related by the equation

$$\partial\phi_b/\partial t = 2\pi r_0^2 k_0 \partial \mathrm{H}_0/\partial t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Assuming that r_0 denotes some fixed radius, a slight adjustment of the bars will change $\partial\phi_b/\partial t$ in relation to $\partial \mathrm{H}_0/\partial t$, and ϕ_b will satisfy an equation

$$\partial\phi_b/\partial t = 2\pi r_0^2 k_0 (1+\eta) \partial \mathrm{H}_0/\partial t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The equilibrium orbit will now be of radius $r_0+\delta r_0$, where, provided η and $(\delta r_0/r_0)$ are both small,

$$\delta r_0/r_0 = k_0\eta/(1-n). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The field in the annular gap is assumed to follow the usual r^{-n} radial law.

To illustrate the orders of magnitude involved in equation (4), values appropriate to the 375 MeV. Glasgow synchrotron are $n=0{\cdot}7$, $r_0=125$ cm., $k_0=0{\cdot}82$. It follows that a misadjustment in the rate of change of ϕ_b in this machine of only 0·3 per cent involves an error in equilibrium orbit location of 1 cm.

(ii) *Magnetic Circuit Equations.*

To express the Wideröe flux condition in practical terms we must consider the physical layout of the synchrotron magnet, *e. g.* as shown in fig. 1. For simplicity of analysis any actual arrangement of betatron bars may be represented nearly enough by a single laminated bar of length L, useful cross section A_b, and carrying a uniform flux density B_b, placed in series with an air gap of length g, area A_g in which the flux density is B_g. Thus for continuity we have

$$\phi_b = A_b B_b = A_g B_g. \quad \ldots\ldots\ldots \quad (5)$$

Let G be the height of the main annular air gap of the synchrotron, at radius r_0, and let H_b denote the magnetic intensity in the betatron bar. Assuming equal return path reluctance for annular and betatron bar fluxes we have

$$g'B_b + LH_b = GH_0 + D, \quad \ldots\ldots\ldots \quad (6)$$

where $g' = gA_b/A_g$ and the term D represents any D.C. bias which may be applied to the betatron bars.

When skin effect in the bar laminations is negligible, B_b and H_b are simply related by a normal D.C. hysteresis loop taken well into saturation. When skin effect is appreciable, B_b becomes the average flux density in a lamination and H_b denotes the intensity at a lamination surface, and some modification of the hysteresis loop is required to give the correct relation. This problem is not considered further here, since the results of § 2 (iii) are not affected, but a method of constructing a suitably modified loop is given later in § 7.

Differentiating equation (6) partially with respect to time, and using equation (5) gives

$$\partial\phi_b/\partial t = (G/R)\partial H_0/\partial t, \quad \ldots\ldots\ldots \quad (7)$$

where R is defined by

$$R = (g/A_g) + (L/sA_b), \quad \ldots\ldots\ldots \quad (8)$$

and s is the "differential permeability" $\partial B_b/\partial H_b$. It follows from equation (2) that an equilibrium orbit of radius r_0 will be established when R satisfies the equation

$$G/R = 2\pi r_0^2 k_0. \quad \ldots\ldots\ldots \quad (9)$$

(iii) *Betatron Air Gap Dimensions.*

To determine the betatron air gap dimensions it is permissible to assume that, in equation (8), $(L/sA_b) \ll g/A_g$ at low flux densities. Equation (9) then approximates to

$$g/A_g = G/(2\pi r_0^2 k_0). \quad \ldots\ldots\ldots \quad (10)$$

For economy g and A_g are made as small as possible, the limiting factor being the tolerance on the dimension g. Equations (4) and (7) show that a small error, δg, in g involves an error in equilibrium orbit location given by

$$\delta r_0/r_0 = -k_0\delta g/(1-n)g. \quad \ldots\ldots\ldots \quad (11)$$

Observations on a 30 MeV. synchrotron (Fry *et al.*, to be published) suggest that, if r_0 denotes the mean radius of the vacuum donut, within 90 per cent of the maximum electron injection into the betatron phase will be obtainable provided δr_0 is less than 0·1 of the radial donut width.

Applying this information, for example, to the 375 MeV. synchrotron, which has $g=1{\cdot}9$ cm. and a donut width of about 12·5 cm., δr_0 must be less than 1·25 cm. From equation (11), g must be correct to 0·007 cm.

Thus equations (10) and (11), plus the requirement for accurate location of the equilibrium orbit at injection, decide the values of g and A_g, and the value of L will generally follow geometrically from these. So far the discussion has involved only a slight restriction on A_b, namely that the approximate equation (10) should be justified. To specify A_b more definitely requires some further specification of betatron bar performance, *i. e.* the rate at which the electron orbits may contract, as a result of bar saturation, while still giving efficient betatron-synchrotron transition.

§3. Betatron-synchrotron Transition.

(i) *Transition Theory.*

The problem of electron trapping at betatron-synchrotron transition has been discussed in several papers on synchrotron theory (*e. g.* Bohm and Foldy 1946), and particular aspects, such as the effect of finite time of rise of the synchrotron R.F. voltage pulse, have been studied (DePackh and Birnbaum 1948; Twiss and Frank 1949). A simple treatment which is sufficiently comprehensive for the purposes of this paper is that given by Goward (1949), who assumes sudden cessation of betatron acceleration and simultaneous application of a steady R.F. voltage. Goward shows that quite different factors determine efficiency of transition in a synchrotron giving say 30 MeV. and one giving a few hundred MeV. To avoid excessive length, the discussion in this section (§3 (i)) is confined to synchrotrons giving from about 100 to 400 MeV., other energy ranges receiving only brief consideration in §3 (ii).

Following Goward's treatment, we assume that electrons are accelerated by betatron action up to a total energy E_s. An R.F. voltage of peak value V is then switched on, trapping a fraction f_1 of the electrons into stable synchrotron orbits. In a high energy machine, Goward shows that V is always large enough to make f_1 approximately unity; *i. e.* nearly all the electrons are trapped. The trapped electrons execute radial oscillations about the equilibrium synchronous orbit (Bohm and Foldy 1946), ranging in amplitude up to a stable limit ρ_s, where ρ_s is a function of V. If V is increased beyond a certain value, V_1, ρ_s exceeds the clearance (ρ_m) between the equilibrium synchronous orbit and the donut wall, and some trapped electrons are consequently lost. V_1 and ρ_m are related by the equation

$$V_1=[(1-n)^2\beta_s^4\pi KE_s/2](\rho_m/r_0)^2, \quad \dots\dots \quad (12)$$

where K is given by

$$K=1+n/[(1-n)\beta_s^2]. \quad \dots\dots\dots \quad (13)$$

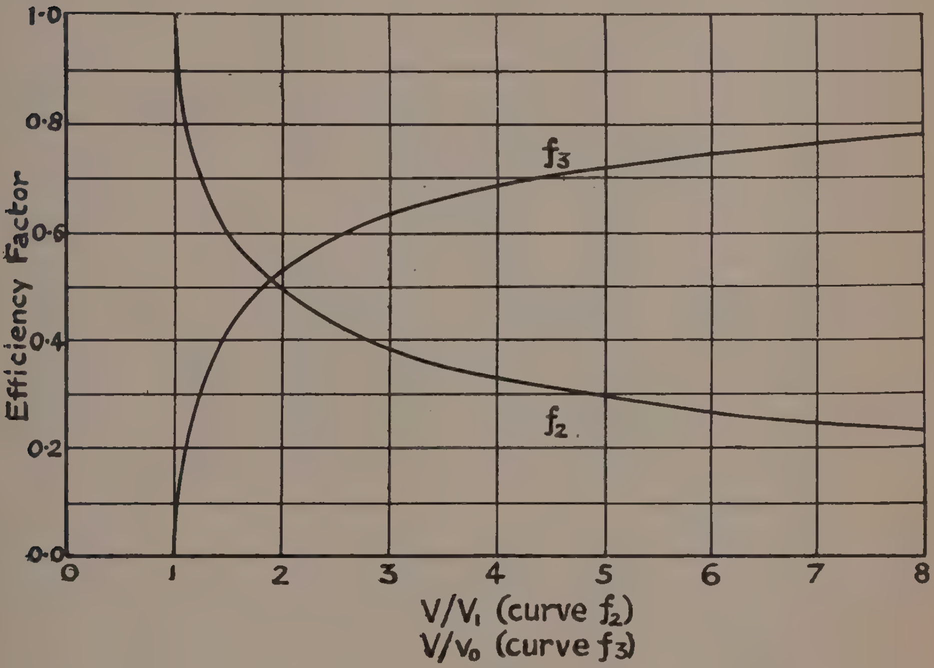
β_s is the electron velocity, at total energy E_s, relative to that of light.

For $V > V_1$ a fraction $(1 - f_2)$ of the electrons collide with the donut walls, while f_2 are accelerated further ; f_2 is given by the equation

$$f_2 = (2/\pi)\sin^{-1}(V_1/V)^{\frac{1}{2}}, \quad \ldots \ldots \quad (14)$$

shown graphically in fig. 2.

Fig. 2.



Electron trapping efficiency factors.
f_2. Result of scatter to donut walls at transition.
f_3. Result of change of stable R.F. phase during synchrotron acceleration.

As synchrotron acceleration proceeds the betatron flux bars saturate and the accelerating " volts per turn ", v, which must be supplied by the R.F. resonator, rise to a maximum v_0. Assuming that electron radiation losses are not of primary importance in synchrotrons of the energy range considered, v_0 is given by the equation

$$v_0 = 4\pi^2 f r_0 E k_0/(3 \times 10^{10}), \quad \ldots \ldots \quad (15)$$

where f is the synchrotron magnet excitation frequency in c/s. and E the final (total) electron energy. Thus $V > v_0$ is a necessary condition for synchrotron acceleration and Goward introduces a further efficiency factor f_3, depending on the ratio V/v_0, which takes into account electrons lost during synchrotron acceleration. The factor f_3 is also shown in fig. 2

and is calculated for high energy machines by neglecting phase damping of the synchrotron oscillations (Bohm and Foldy 1946). This damping is, of course, important, but is difficult to include in a sufficiently general manner. To a sufficient approximation, the fraction of original betatron electrons accelerated to peak energy is given by f_2 or f_3, whichever is the least, and hence the most efficient value of V to use is one which makes $f_2=f_3$, *e.g.* for 65 per cent efficiency, putting $f_2=f_3=0{\cdot}65$, fig. 2 gives $(V/v_0)=3{\cdot}3$ and $V/V_1=1{\cdot}35$. Thus we have

$$V_1=2{\cdot}45\,v_0. \qquad (16)$$

By means of an example, it is shown in the next section that an equation such as (16) forms a convenient basis for specifying betatron flux bar performance. It must be emphasized, however, that the equation is based on essentially pessimistic calculations of electron trapping efficiency. Even a simple rectangular R.F. pulse has in practice an appreciable time of rise, and phase oscillation damping will reduce loss of electrons to the walls. The same damping appreciably reduces loss during synchrotron acceleration, and thus both f_2 and f_3 tend to become nearer unity. Moreover it is possible to reduce both kinds of loss by suitable shaping of the R.F. pulse (DePackh and Birnbaum 1948; Twiss and Frank 1949). A relatively small R.F. voltage would suffice for initial trapping of the electrons ($f_1\simeq 1$), after which, to allow for the increasing volts per turn v, the applied R.F. voltage would have to be increased. This latter process in itself gives increased oscillation damping.

(ii) *Flux Bar Performance.*

Taking the 375 MeV. synchrotron as example, the parameters in equation (15) are $f=50$ c/s., $r_0=125$ cm., $E=375$ MeV., $k_0=0{\cdot}82$. Hence $v_0=2{\cdot}5$ kV., and from equation (16) the condition for 65 per cent efficient overall electron trapping is $V_1=6{\cdot}1$ kV. In equation (12), however, $n=0{\cdot}7$, and assuming $\beta_s\simeq 1$ gives

$$V_1=(3/10^5)E_s\rho_m^2. \qquad (17)$$

Since $V_1=6{\cdot}1$ kV. we have

$$E_s\rho_m^2=203\times 10^6. \qquad (18)$$

Equation (18) defines pairs of values of E_s and ρ_m which permit 65 per cent efficient transition, on the basis of the above simple theory. Since the effective donut width is 12·5 cm., if we write $(6{\cdot}25-\rho_m)=\Delta r$, the distance Δr represents the permissible radial displacement of the betatron beam centre from r_0, at transition energy E_s (assuming that transition occurs when betatron beam and stable synchronous orbit coincide). A table of values of E_s and Δr may now be drawn up, as in Table I.

TABLE I.

E_s(MeV.)	5·2	6·2	7·4	9·0	11·3
Δr(cm.)	0	0·5	1·0	1·5	2·0

Thus in addition to establishing an equilibrium orbit of radius r_0, when initial electron injection into the betatron occurs, the betatron bars must satisfy the additional requirement of maintaining the electron orbits constant within the limits of Table I. The method of extending the bar design to meet such a requirement is described in the following sections.

It is convenient at this point to consider briefly how the discussion in this section (§3) must be modified for synchrotrons of other energies. Above 400 MeV., electron radiation loss becomes important, and may determine the maximum R.F. " volts per turn " v_0 (see Appendix II). Moreover R.F. pulse shaping becomes essential to avoid low f_2 values (DePackh and Birnbaum 1948) and the values of E_s and Δr will be governed by a more complex transition theory than that of Goward (1949).

Considering energies below 100 MeV., the values of v_0 and of V become smaller, and loss to the donut walls at transition (f_2) becomes less likely. The value of Δr is not so critical, although Δr must always be less than half the donut width (*i. e.* the stable synchronous orbit and betatron beam must always coincide within the donut limits). The factor f_1, however, which depends on E_s and V, now becomes more important than f_2 or f_3, and the design problem consists essentially of arranging a high enough value of E_s to make $f_1 \simeq 1$, taking into account the value of V available. A full treatment is given in Goward's paper.

§4. BETATRON FLUX BAR CROSS SECTION.

(i) *Instantaneous Circle Orbit Equations.*

Neglecting free oscillations, individual electrons in a betatron beam move on so-called " instantaneous circle " orbits (Kerst and Serber 1941) which, at any time, have a range of radii corresponding to the energy spread of the electrons. It is convenient, for the following discussion, to consider the electron motion in terms of these instantaneous circle orbits, and to use the instantaneous circle orbit equation (Amaldi and Ferretti 1946):

$$\phi_r = 2\pi r^2 H_r + C. \qquad \qquad (19)$$

H_r is the vertical field at radius r and ϕ_r is the total flux linking the circle of radius r. If C is a constant, the equation defines the variation of r for any electron throughout its betatron motion.

For design purposes it is sufficient to consider an electron injected from the electron gun so that, initially, $r = r_0$. (This electron roughly defines the mean orbit radius of the whole electron beam.) At injection, equation (19) for this electron takes the form

$$\phi_0 = 2\pi r_0^2 H_0 + C. \qquad \qquad (20)$$

After some betatron acceleration, r will in general have changed to some value $r_0 + \Delta r$, where we assume $(\Delta r / r_0)$ is small. Corresponding to this, the flux through the circle r_0 will then be given by

$$\phi_0 = 2\pi r_0^2 H_0 + C + \Delta\phi_0, \qquad \qquad (21)$$

where $(\Delta\phi_0/\phi_0)$ is also small. $\Delta\phi_0$ and Δr are related by the equation

$$\Delta\phi_0/(2\pi r_0^2 H_0)=(1-n)(\Delta r/r_0). \qquad (22)$$

(ii) *Determination of Betatron Bar Cross Section.*

From §2 the betatron air gap dimension and bar length L (as defined in §2) are already determined. For the present purpose, H_0 is nearly enough zero when injection into the betatron occurs. Thus when $H_0=0$, let the variables in equation (6) have the values $B_b=B_1$, $H_b=H_1$. Then for any value of H_0 we have

$$B_b-B_1=(G/g')H_0-(L/g')(H_b-H_1). \qquad (23)$$

Equation (23) gives the total change in B_b corresponding to an increase in field at r_0 from zero to H_0.

A theoretically ideal betatron bar would have constant differential permeability s, and would maintain the electron considered in §4(i) at a constant orbit radius $r=r_0$. Let the parameter R in equation (7) have the value R_1 when $B_b=B_1$ (and $s=s_1$). Then, for an ideal bar, R would be constant at R_1, and equation (7) gives

$$B_b-B_1=GH_0/(A_bR_1). \qquad (24)$$

Comparison of (23) and (24) gives the amount, ΔB_b, by which the actual bar density differs from the ideal value, as a result of variation of s and R. Assuming all non-linear effects to be confined to the betatron bar flux, we may identify $A_b\Delta B_b$ with the $\Delta\phi_0$ of equation (22), and obtain

$$(R_1/L)\Delta\phi_0=(B_b-B_1)(1/s_1)-(H_b-H_1). \qquad (25)$$

If we neglect skin effect for the moment, the right hand side of equation (25) has a simple graphical meaning in terms of the D.C hysteresis loop as shown in fig. 3, and may conveniently be denoted by $(-\Delta H_b)$. Using equations (9) and (22), we have

$$-\Delta H_b=[(1-n)G/Lr_0k_0]H_0\Delta r. \qquad (26)$$

Further, using the well known relation

$$E_s\beta_s=300r_0H_0, \qquad (27)$$

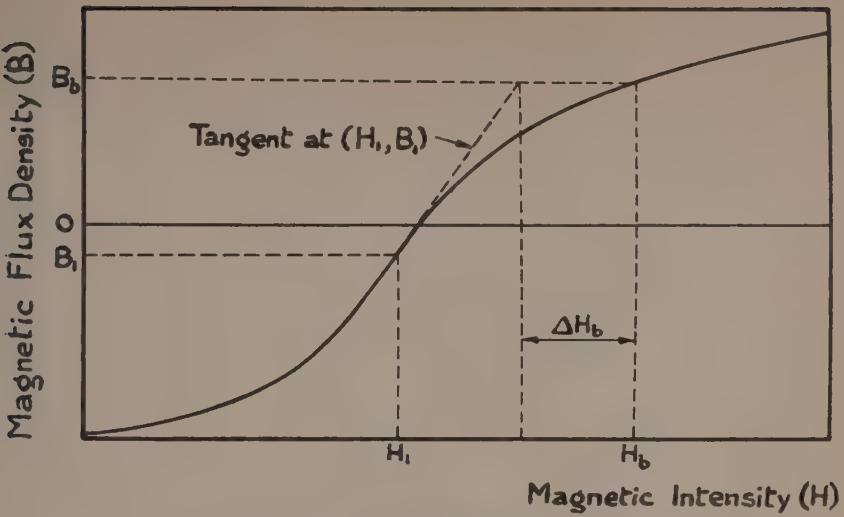
equation (26) may alternatively be written

$$-\Delta H_b=[(1-n)G/(300\,Lr_0^2k_0)]E_s\beta_s\Delta r. \qquad (28)$$

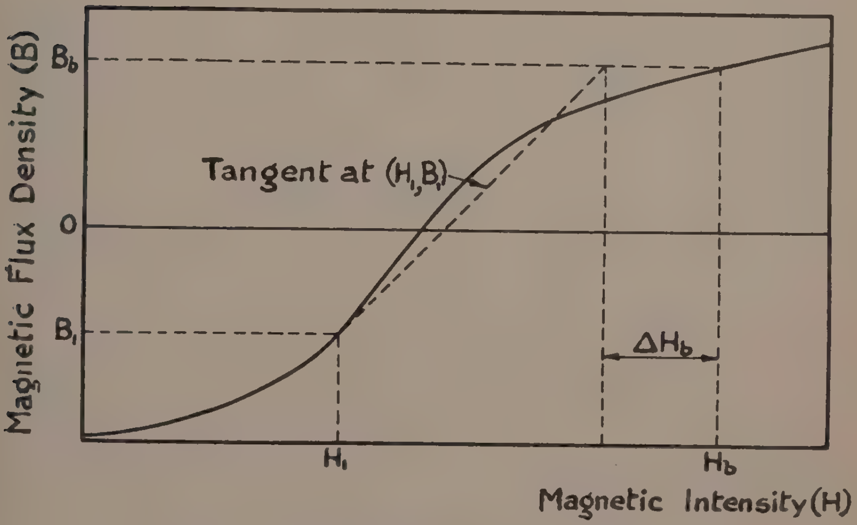
In conjunction with a pair of values of E_s and Δr, specified for example as in §3, equation (28) effectively determines a betatron bar cross section. The derived value of ΔH_b specifies pairs of values of B_b and B_1 from the hysteresis loop of the bar material, as in fig. 3. From equations (9), (24) and (27), A_b is then given in terms of such a pair of values by the equation

$$A_b=2\pi r_0E_s\beta_sk_0/300(B_b-B_1). \qquad (29)$$

Fig. 3.



(a) Little (or no) Negative Flux Bias.



(b) Large Negative Flux Bias.

Hysteresis loop construction for determining changes in electron orbit radius.

In the following section, the application of the above procedure is illustrated by a specific example.

§5. EXAMPLE DESIGN : 375 MEV. SYNCHROTRON.

(i) *No Applied D.C. Flux Bias.*

Taking again the example of the 375 MeV. Glasgow synchrotron, the known parameters in equation (28) are $n=0{\cdot}7$; $G=10$ cm., $L=100$ cm., $r_0=125$ cm. and $k_0=0{\cdot}82$. If at least 65 per cent efficient transition is taken as a design criterion, Table I. gives appropriate values of E_s and Δr. For example, the values $E_s=9{\cdot}0$ MeV., $\Delta r=1{\cdot}5$ cm., inserted in equation (28), give $-\Delta H_b=0{\cdot}105$. Drawing a tangent at the point $B_1=0$ to the hysteresis loop of the transformer steel laminations to be used yields the value $B_b=4600$ gauss, in the manner indicated in fig. 3 (*a*). From equation (29), $A_b=4200$ sq. cm.

Other values of E_s and Δr from Table I. give cross sections greater than 4200 sq. cm. Hence we conclude that if no D.C. bias is to be used, and skin effect in the laminations is negligible, a cross section of 4200 sq. cm. is the minimum which will ensure at least 65 per cent transition efficiency.

(ii) *Application of D.C. Bias.*

An auxiliary D.C. winding on the betatron bars may be used to give B_1 some value other than zero, and to effect a saving in bar cross section. The graphical construction may now take the form of either fig. 3 (*a*) or 3 (*b*), according to the amount of bias used. 2000 gauss negative bias, for example, gives a fig. 3 (*a*) type of construction, and the same design procedure as in §5 (i) yields a cross section of 3200 sq. cm. A greater bias gives the type of construction shown in fig. 3 (*b*). Here, as is readily seen from equation (26) or (28), the electron orbit, initially of radius r_0, will first expand to some extent, and then contract through r_0 as tangent and hysteresis loop intersect. The design procedure is unaltered, and taking B_1 as 4000 gauss negative gives a cross section of 2000 sq. cm., less than half that determined for zero bias.

Equation (26) may be used to check that the initial orbit expansion under fig. 3 (*b*) conditions is not excessive ; *e.g.* in the above 4000 gauss bias example, instantaneous circle orbit expansions from 125 cm. at injection to a maximum of 126 cm. at about 3·6 MeV. electron (total) energy will occur.

By the above methods a betatron bar cross section can be determined to give any specific degree of electron orbit contraction. One single value of cross section is obtained if no flux bias is to be used, but a variety of combinations of cross section and bias will give essentially the same results. So far the question of skin effect in the bar laminations has been neglected and we now proceed to determine whether practical lamination thicknesses involve any serious modifications in betatron bar performance.

§6. SKIN EFFECT.

(i) *Modification of Hysteresis Loop.*

The use of practical lamination thicknesses in betatron bar construction will in general involve appreciable skin effect, and the variables B_b and H_b

used in §§4 and 5, should strictly be defined respectively as the average flux density in a betatron bar lamination and the intensity at a lamination surface. Some modified hysteresis loop is therefore required to relate B_b and H_b if the graphical method of §4 is to yield a correct result.

Consider a lamination extending in thickness between the limits $-\xi/2 \leqslant x \leqslant \xi/2$. The flux density, B, and intensity, H, at any point in the lamination (directed say along the y axis) are considered to be functions of x and time, t, only, and are related by the well known equation

$$\partial^2 H/\partial x^2 = (4\pi/10^9\rho)\partial B/\partial t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where ρ is the electrical resistivity of the material. Solutions of the equation when B is a linear or sinusoidal function of t are also well known, but to a first approximation the particular function of time is immaterial, and it is found that $\partial B/\partial t$ may be treated as a constant with respect to x. Equation (30) is then simply integrable, and it follows that the distribution of H is of the form $(a+bx^2)$. If H_b and H_c are the values of H when $x=\pm\xi/2$ and 0 respectively, then

$$H_b - H_c = (\pi/10^9 . 2\rho)\xi^2 \partial B_b/\partial t, \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where B_b is the value of B averaged over x.

Assuming that the synchrotron magnet excitation is derived from a resonant circuit of frequency f, the magnetic field H_0 will not vary sinusoidally, since saturable betatron bars introduce non-linearity into the circuit. If the ratios of magnet inductances with unsaturated and saturated bars is q, then during the initial betatron phase

$$\partial H_0/\partial t = 2\pi f H_m/q, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where H_m is the peak value of H_0. Using equations (7), (9), (27) and (31) we obtain

$$H_b - H_c = (2{\cdot}07/10^{10}) r_0 k_0 f E \xi^2/q A_b \rho, \quad . \quad . \quad . \quad . \quad (33)$$

where E is the peak electron energy (eV.).

Corresponding to the range of variation of H given by equation (33), the flux density B will vary from B_s at the lamination surface to B_c at its centre, where (H_b, B_s) and (H_c, B_c) are points on the normal D.C. hysteresis loop. If we approximate to this range of the hysteresis loop by a straight line, the distribution of B will also be of the form $(a+bx^2)$, and by simple integration

$$B_b - B_c = (B_s - B_c)/3. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

As indicated in fig. 4 this approximate result gives a graphical method for deriving from the D.C. hysteresis loop an (H_b, B_b) curve representing true operating conditions.

(ii) *Example: 375 MeV. Synchrotron.*

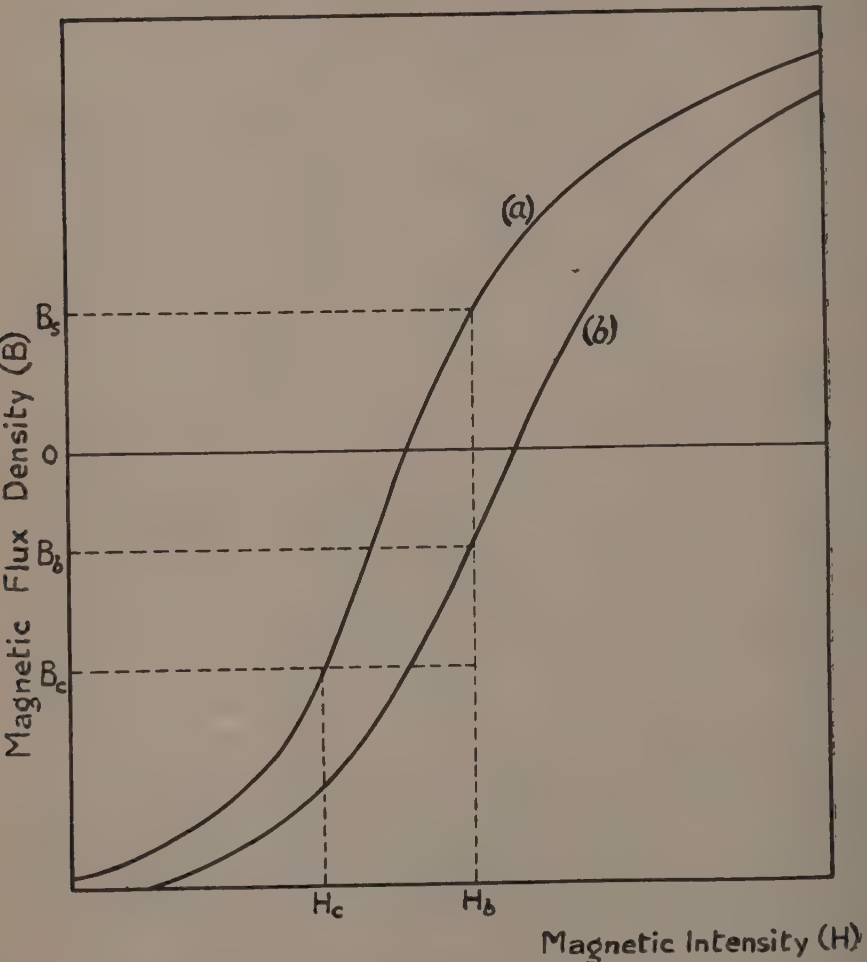
Again taking a specific example, the 375 MeV. synchrotron, the proposed bar material consists of 0·007 in. laminations of resistivity

50 microhms per cm. cube. With $r_0=125$ cm., $k_0=0{\cdot}82$, $f=50$ c/s. and $q=3$, equation (33) gives

$$H_b-H_c=830/A_b, \qquad (35)$$

if A_b is expressed in sq. cm. By neglecting skin effect, a cross section of 4200 sq. cm. was determined as suitable for use with no D.C. bias.

Fig. 4.



Modification of hysteresis loop by skin effect.
(*a*) D.C. hysteresis loop,
(*b*) Modified loop $B_b-B_c=(B_s-B_c)/3$.

For this cross section $(H_b-H_c)=0{\cdot}2$, which is small enough for the approximate methods of the last section to be applied. Deriving the modified loop, as in fig. 4, and applying the construction of fig. 3 (*a*), with $B_1=0$, indicates orbit contractions slightly less than those predicted by the unmodified loop. Neglecting skin effect resulted in an over estimation of the cross section required by only a few per cent.

Similar calculations may be made on the assumption of A_b values determined in §5(ii). The practical result is to predict that, for a given cross section, rather less bias would be required than that specified in §5(ii). For example a 2000 sq. cm. cross section would require about 3000 rather than 4000 gauss negative bias. The detailed variations in electron orbit radius are of course affected both by the change of bias and the modified loop shape, and rechecking for excessive initial expansion as in §5(ii), shows expansion up to 127 cm. at 5 MeV. total electron energy.

It is reasonable to conclude that in general, provided equation (33) indicates no excessive skin effect, use of a normal D.C. hysteresis loop is sufficiently accurate for initial design purposes. A final design should, however, be checked by the appropriate modified loop.

§7. Conclusion.

The essential steps in the magnetic design of betatron bars for an electron synchrotron may now be summarized as follows. Betatron air gap dimensions are determined from equations (10) and (11), and the flux bar length generally follows simply from geometrical considerations. The methods of §3 define the rate of radial shrinkage of the electron beam orbit which may be permitted if efficient betatron-synchrotron transition is to be ensured. Equation (28) essentially expresses orbit shrinkage rate in terms of non-linearity of the betatron bar material, and a graphical construction on the hysteresis loop, as shown in fig. 3, together with equation (29) gives the appropriate cross sectional area of bar. As a check on the validity of neglecting skin effects equation (33) should be evaluated, and a final design should be checked by proceeding to construct a modified hysteresis loop by the method of §6(i) and fig. 4.

Acknowledgments.

Acknowledgment is made to many helpful suggestions made to the author at Metropolitan Vickers Electrical Company, by members of the Glasgow Synchrotron Committee. This paper was written as part of the accelerator programme of the Electronics Division of the Atomic Energy Research Establishment, and is published with the permission of the Director.

References.

Amaldi, E., and Ferretti, B., 1946, *Rev. Sci. Instrum.*, **17**, 389.
Blewett, J. P., 1945, *Phys. Rev.*, **69**, 87.
Bohm, D., and Foldy, L., 1946, *Phys. Rev.*, **70**, 249.
DePackh, D. C., and Birnbaum, M., 1948, *J. Appl. Phys.*, **19**, 795.
Goward, F. K., 1949, *Proc. Phys. Soc.* A, **62**, 617.
Kerst, D. W., and Serber, R., 1941, *Phys. Rev.*, **50**, 53.
Kerst, D. W., 1945, *Phys. Rev.*, **69**, 233.
McMillan, E. M., 1945, *Phys. Rev.*, **68**, 144.
Twiss, R. Q., and Frank, N. H., 1949, *Rev. Sci. Instrum.*, **20**, 1.

APPENDIX I.

ELECTRICAL CONTROL AND CORRECTION OF BETATRON FLUX.

The effect of applying D.C. bias to betatron-starting bars has already been considered in the main paper, and it is shown that considerable improvement in performance may be obtained. Other methods of counteracting non-linearity in the betatron bars by auxiliary electrical circuits are available, and the techniques involved are being developed at Metropolitan-Vickers Electrical Co., Ltd., in association with the 375 MeV. synchrotron magnet design. However, a brief review may be appropriate to this paper.

One possible general technique is suggested by Kerst's (1945) flux forcing circuit, which consists of an auxiliary winding around main poles and betatron bars, automatically ensuring the correct relation between annular and central fluxes. In application to a synchrotron the circuit must be "switched off" after betatron-synchrotron transition has occurred, to avoid excessive currents. Mechanical or electronic switching are possible solutions, if care is taken over transient effects. A more elegant modification is an electronic circuit in which, for example, any unbalanced voltage from a Kerst type winding is fed into an amplifier which in turn feeds a correcting winding on the betatron bars. Switching should then be relatively simple.

Electrical control may also be used to give fine adjustment of the equilibrium orbit radius, an essential operational facility. Mechanical adjustment of the betatron air gap alone may be neither convenient nor sufficiently sensitive. (See for example, the air gap tolerance figures of §2 (iii)). An auxiliary bar winding connected across a variable inductance conveniently achieves the desired result.

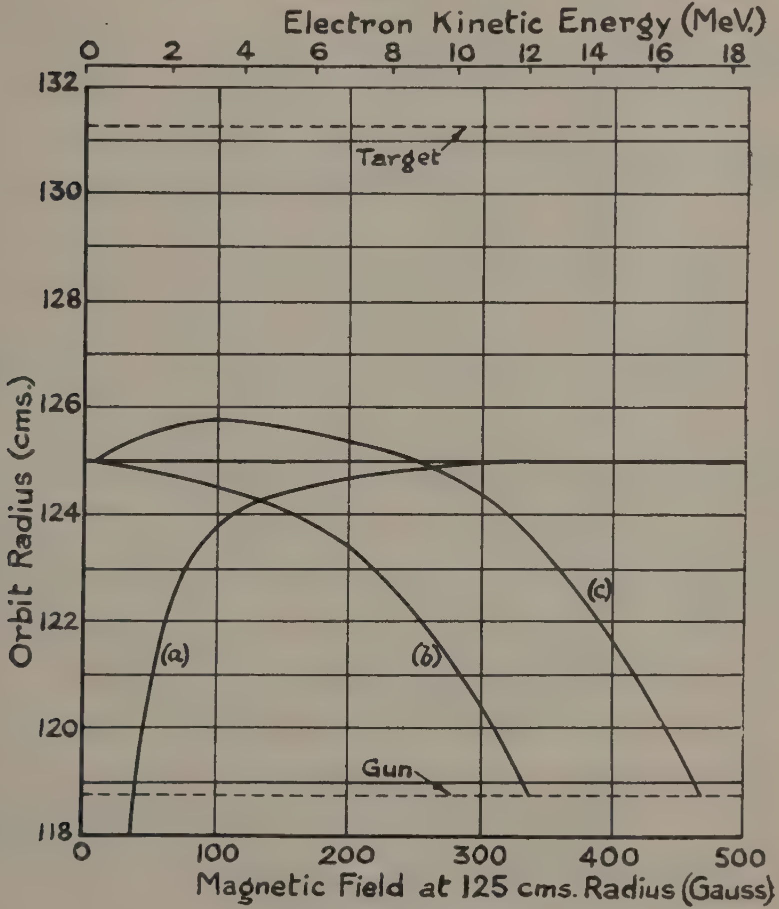
APPENDIX II

375 MeV. SYNCHROTRON PERFORMANCE CHARACTERISTICS.

In the main paper the design methods discussed were illustrated at various points by calculations based on the 375 MeV. synchrotron, and in particular it was determined (§§3 (ii) and 5 (i)) that a betatron bar of 100 cm. effective length and 4200 sq. cm. cross section, of the specified material and used without D.C. bias, would make at least 65 per cent betatron-synchrotron transition possible. In the specification finally adopted for the machine, the effective flux bar dimensions are 100 cm. length and 3300 sq. cm. cross section, and using this specification it was considered worth while to calculate the expected performance in some detail. Some of the more instructive results of the calculations are presented below. Since the machine will initially be operated at 300 MeV., some results for this energy are included.

Using a hysteresis loop corrected for skin effect by the method of §6, the instantaneous circle motion of an electron, initially injected on to the equilibrium orbit, was calculated for various flux bar D.C. bias values, by means of equation (26). The difference in skin effect for 300 and 375 MeV. operation is too small to affect the results appreciably. Fig. 5 shows the calculated orbits, and a superposed curve shows the synchronous orbit radius expanding towards r_0 as the electron velocity approaches that of light. Betatron-synchrotron transition takes place at the intersection of this curve with a betatron orbit curve, and it would appear that transition conditions may be materially improved by flux biasing or other electrical correction methods. Further adjustment may be acheived by alteration

Fig. 5.



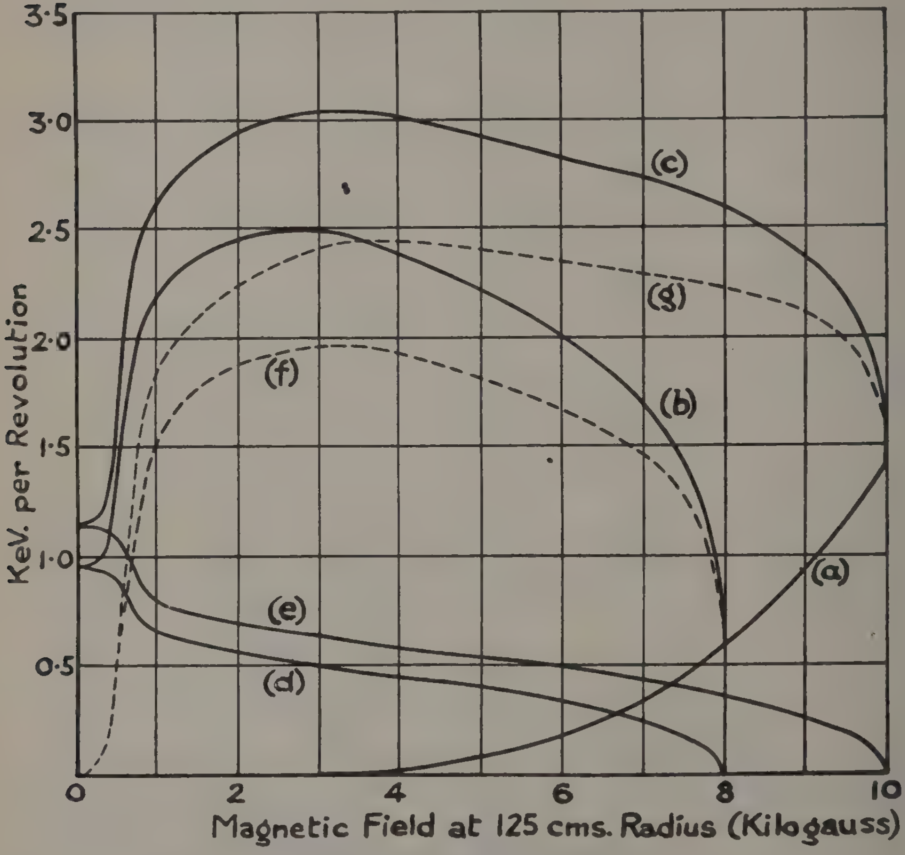
Variation of orbit radii.
375 MeV. synchrotron.

(*a*) Synchronous orbit.
(*b*) Betatron electron orbit. Zero betatron bar flux bias.
(*c*) Betatron electron orbit. 2000 gauss betatron bar flux bias.

of the resonator frequency, making the synchronous radius tend to some value other than 125 cm.

The betatron output energy of the machine may also be deduced from fig. 5; *e. g.* with no flux biasing, electrons would strike the internal gun at about 12 MeV. energy. (An external target and internal gun are to be used in this machine.)

Fig. 6.



Volts per revolution diagram.
375 MeV. synchrotron.

(*a*) Incoherent radiation loss.
(*b*), (*c*) Total volts per revolution required for acceleration to (*b*) 300 MeV., (*c*) 375 MeV.
(*d*), (*e*) Volts per revolution supplied by betatron effect, (*d*) 300 MeV., (*c*) 375 MeV.
(*f*), (*g*) Volts per revolution derived from R.F. cavity, (*f*) 300 MeV., (*g*) 375 MeV.

To maintain a constant orbit radius the electron energy must increase by $(2\pi r_0^2/10^8)\partial H_0/\partial t$ electron volts per revolution. Incoherent radiation of energy by the electrons (MacMillan 1945) is equivalent to an energy loss (also in eV. per revolution) given by the expression $(6/10^7 r_0)(E_t/m_0c^2)^4\beta^3$, where E_t is the total electron energy and m_0c^2 its rest energy. Some coherent radiation loss also occurs, but assuming that the total charge associated with the electron beam will be about 1/100 (or less) of the theoretical maximum value 0·05 microcoulomb (Blewett 1945), this loss is likely to be comparatively small. Thus betatron effect and the R.F. voltage must supply to the electrons an energy represented by the sum of the two above expressions.

Fig. 6 is an "eV. per revolution diagram" showing the relative magnitudes of the above effects. The R.F. volts per turn (v) rises relatively slowly (adiabatically) to the maximum value v_0. Synchrotron phase oscillation damping will considerably reduce electron loss during this period, giving more efficient overall electron trapping than that indicated by Goward's (1949) methods. It is thus of interest to attempt an approximate estimate of the overall efficiency, taking some account of phase damping, but continuing to assume a simple rectangular R.F. voltage wave form.

Bohm and Foldy (1946) and several other authors have discussed the phase oscillations of electrons in a synchrotron about the stable R.F. phase ψ, where ψ is given by

$$\sin\psi = v/V. \qquad (1)$$

When ψ is zero, stable oscillations are possible up to a full 180° amplitude, but as ψ increases from zero, the permissible amplitude decreases and, moreover, the oscillations become assymetric about ψ. Thus, when ψ is 30° the maximum stable oscillation is from $-40°$ to $+150°$, which is $-70°$ to $+120°$ about ψ. In order to make the following trapping efficiency calculations pessimistic rather than optimistic, we consider only the smaller phase amplitude (such as 70° when ψ is 30°), and denote this by θ_m. At all times during synchrotron acceleration the electron bunch must contain a range of smaller phase amplitudes from zero up to θ, where $\theta \leqslant \theta_m$.

At transition energy, v and ψ are approximately zero and θ_m is 180°. If some loss to the walls occurs at transition, then from the definition of f_2 in equation (14), the remaining electrons have phase amplitudes up to only (180° f_2). After some synchrotron acceleration has taken place, and ψ is no longer zero, all electrons which tend to have phase amplitudes greater than θ_m are considered lost. Thus, neglecting damping for the moment, of the electrons remaining in the beam immediately after transition a fraction u will now remain, where (if θ_m is expressed in degrees)

$$u = \theta_m / 180 f_2. \qquad (2)$$

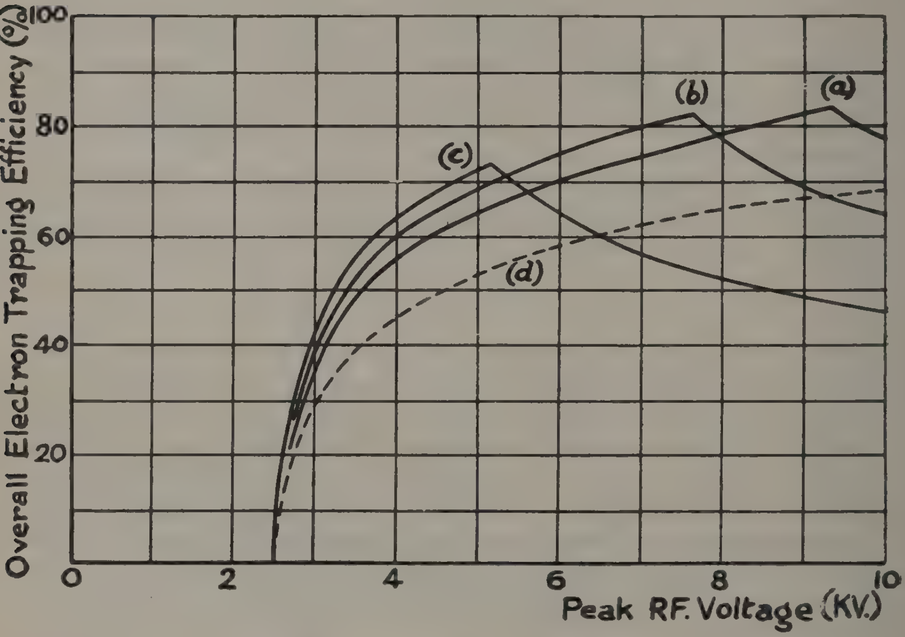
This is essentially the calculation used by Goward (1949).

It has been shown (*e. g.* Bohm and Foldy 1946) that synchrotron phase oscillations of small amplitude are damped in proportion to $(E_t \cos\psi)^{\frac{1}{4}}$. The phase oscillation equation can only be solved for all amplitudes when

ψ is zero (*e. g.* Twiss and Frank 1949), and, from the form of the resulting equation, the rate of damping appears to increase with the amplitude, approximating to $E_t^{\frac{1}{4}}$ over a wide amplitude range. Thus if the small amplitude damping law is assumed to hold for all amplitudes and ψ values, the tendency again should be to a pessimistic rather than optimistic result. Equation (2) is accordingly modified to the form.

$$u=(\theta_m/180f_2)(E_t \cos\psi/E_s)^{\frac{1}{4}}. \qquad (3)$$

Fig. 7.



Overall electron trapping efficiency.
375 MeV. synchrotron.

(*a*) Transition at 8·0 MeV. total electron energy.
(*b*) Transition at 6·5 MeV. total electron energy.
(*c*) Transition at 5·0 MeV. total electron energy.
(At 0·75 cm. off the donut centre).
(*d*) Efficiency factor f_3 (neglecting phase oscillation damping).

Fig. 7 shows overall trapping efficiency curves, determined by the data of fig. 6 and equation (3). Curves (*a*) and (*b*) assume transition exactly at the donut centre, while curve (*c*) represents the uncorrected transition of fig. 5, for which the betatron beam is displaced by 0·75 cm., from the donut centre. For convenience of comparison curve (*d*) has been added, representing the efficiency factor f_3 of fig. 2 and showing the effect of neglecting electron oscillation damping.

As in Goward's calculations, the rising part of an efficiency curve represents conditions under which loss of electrons during the synchrotron phase is predominant, and the falling part is determined by the loss at transition. It appears that an R.F. voltage of about 5 kV. peak will ensure 70 per cent overall trapping efficiency, without flux correction, and higher efficiencies are possible under adjusted conditions. Moreover, the effect of finite R.F. switch on time, which has been neglected, will be to give additional efficiency.

IV. *On the Interaction of Colloidal Particles.*—III. *Some Theorems Concerning the Free Energy of Double Layers.*

By S. Levine, Ph.D.,
Birkbeck College Research Laboratory*.

[Received July 26, 1949.]

Summary.

Making use of the Poisson-Boltzmann equation, a number of theorems relating to the free energy of the electric double layers of colloidal particles in dilute sols are developed. A general proof is given for the equivalence of two methods for determining this free energy, viz. by (i) using the fictitious process of charging the excess ions in the double layers subject to the condition of constant surface density of ions and (ii) imagining that the particles are charged by bringing ions from the interior of the dispersion medium to the surface. In the case of two infinite parallel plates, the difficulty that arises due to the infinity of the surface potential, when all the ions are supposed completely discharged, is satisfactorily explained.

The present work is an extension of some of the results obtained by Verwey and Overbeek in their recent book on the theory of the stability of lyophobic colloids. Our theorem of the equivalence of the two methods of evaluating the free energy is more general than a similar theorem proved by Casimir in this book. We are able to derive the Casimir theorem by introducing special conditions.

§1. Introduction.

In their recent book, Verwey and Overbeek (1948) develop two methods of evaluating the free energy associated with the double layers of charged colloidal particles, namely by (i) charging the excess ions in the diffuse part of the double layers under the condition of constant surface potential and (ii) using a generalized Lippmann equation. A general proof of the equivalence of these two methods is given in the book by Casimir (1948).

---

* Communicated by the Author.

In this paper we derive a corresponding theorem, where the distribution of surface ions, and not the surface potential, is kept fixed, by means of an argument similar to that used by Casimir. Our equivalence theorem can be derived without the use of the Poisson-Boltzmann equation (Levine 1949 b) but the proof given here does depend on this equation (or rather a generalized version of it) and; in addition, only applies to dilute sols. Our theorem is more general than that proved by Casimir, since it is valid if the surface distribution of ions is arbitrary (but unchanged during the charging process) and, furthermore, Casimir's theorem can be deduced by assuming that the surface potential be constant during charging.

Verwey and Overbeek apparently did not investigate our method of charging the ions in any detail because of the anomaly of an infinite potential (for the discharged state) in the case of two infinite parallel plates. This difficulty was also encountered by the author and, in fact, lead to an erroneous conclusion (Levine 1946) concerning the nature of the forces between the plates in dilute sols (an attraction was obtained at large separations, whereas the correct result is a repulsion). However, in § 3, we are able to overcome this obstacle in a satisfactory manner.

§ 2. ENERGY OF DOUBLE LAYER OF SINGLE PARTICLE.

For the sake of simplicity we begin with a single colloidal particle of any shape immersed in a large volume of electrolyte. Let the surface charge arise from the adsorption of ions of type 1, charge e_1: ν is the density of such ions at position $\mathbf{s}$ on the surface. If n is the total number of adsorbed ions, then $n=\int_S \nu\, d\mathrm{S}$, integrating over the surface (area) S of the particle. We imagine that all the ions in the system, including the adsorbed ones, are subject to the familiar Debye-Huckel charging process. Assuming that all the ions are being charged at the same rate, the fraction λ will denote the stage in this fictitious process. Our problem is to determine the free energy of the double layer at any stage $\lambda' \leqslant 1$. Now in an earlier paper (Levine 1946), we have already remarked that our method of determining this free energy is based on keeping the surface distribution of ions fixed when increasing λ from 0 to λ'. Thus ν does not depend on any intermediate value $\lambda < \lambda'$. It is also assumed that the distribution ν can be chosen arbitrarily, *i.e.* we treat it as a thermodynamic variable. There is really no loss in generality if we assume that ν has the form

$$\nu = n\nu_0(\lambda', \mathbf{s}), \qquad \int_S \nu_0(\lambda', \mathbf{s})\, d\mathrm{S} = 1, \quad \ldots\ldots \quad (1)$$

where $\nu_0(\lambda', \mathbf{s})$ is subject to the normalizing condition, but it is otherwise an arbitrary function of λ' (which represents here the completion of the charging process) and of the surface position $\mathbf{s}$. In this way n takes on the simple rôle of a " magnitude " factor.

The density of ions of species $i (i=1, 2, \ldots, s)$ in the bulk of the solution is assumed constant, independent of n, and will be denoted by n_i^0.

The average potential at position **r** in the diffuse layer of the particle is ψ, which is taken to be the same for the different types of ions. Casimir (1948) writes the Poisson equation for ψ in the form

$$\nabla^2\psi = -4\pi\rho/D = \lambda f(\lambda\psi), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where ρ is the average charge density in the diffuse layer, D is the dielectric constant of the dispersion medium and f is an arbitrary function. This becomes the Poisson-Boltzmann equation if we put

$$f(\lambda\psi) = -\frac{4\pi}{D}\sum_{i=1}^{s} n_i^0 e_i \exp(-\lambda e_i\psi/kT), \quad . \quad . \quad . \quad . \quad (3)$$

where e_i is the (full) charge of an i ion, k is Boltzmann's constant and T is the temperature. A different form for $f(\lambda\psi)$ has been chosen by Grimley and Mott (1947) when they take into consideration the hydration shell about each ion. The general conclusions in this paper do not depend on the particular form of the function $f(\lambda\psi)$. The average electrostatic energy associated with the excess charge in the double layer at stage λ is

$$E(\lambda, n\,;\ \lambda') = \tfrac{1}{2}\lambda e_1 \int_S \psi\nu\, dS + \tfrac{1}{2}\int_V \psi\rho\, dv. \quad . \quad . \quad . \quad . \quad (4)$$

The second integral in (4) is taken over the volume V of the outer diffuse layer, which may be assumed to extend to infinity, where the potential ψ is zero. Then the corresponding electrical free energy at stage λ' is given by (see Levine 1948)

$$F_e(\lambda', n) = 2\int_0^{\lambda'} \frac{1}{\lambda} E(\lambda, n\,;\ \lambda')\, d\lambda. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The quantities ψ and ρ depend on λ, λ', ν (and hence n) and the coordinates **r** or **s** but to simplify the notation, we shall not indicate this explicitly. It will be clear from the text whether we are referring to the potential at position **r** in the outer layer or at position **s** on the surface.

It is convenient to evaluate the quantity

$$\frac{\partial F_e(\lambda', n)}{\partial n} = 2\int_0^{\lambda'} \frac{1}{\lambda}\frac{\partial E(\lambda, n\,;\ \lambda')}{\partial n}\, d\lambda. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Since we keep ν and hence n fixed when integrating with respect to λ, n and λ are considered independent of one another, and we can differentiate under the integral sign with respect to n. The partial derivatives of the quantities ψ and ν with respect to λ or n will be denoted by a subscript; we note that $\nu_\lambda = 0$. Substituting (2) for ρ into (4), we readily obtain

$$\frac{\partial E(\lambda, n\,;\ \lambda')}{\partial n} = \tfrac{1}{2}\lambda e_1 \int_S (\psi_n \nu + \psi\nu_n)\, dS - \frac{D\lambda}{8\pi}\int_V (\psi_n f + \lambda\psi\psi_n f')\, dv, \quad . \quad (7)$$

where f' if the derivative of f with respect to the argument $\lambda\psi$. Now from (2) we have the relations

$$\nabla^2\psi_\lambda = f + \lambda(\psi + \lambda\psi_\lambda)f', \qquad \nabla^2\psi_n = \lambda^2\psi_n f'. \quad . \quad . \quad . \quad . \quad (8)$$

Then use of (2) and (8) and application of Green's theorem yields the following formulæ for the volume integral on the right of (7)

$$\frac{D\lambda}{8\pi}\int_V (\psi_\lambda\nabla^2\psi_n-\psi_n\nabla^2\psi_\lambda)\,dv=-\frac{D\lambda}{8\pi}\int_S\left(\psi_\lambda\frac{\partial\psi_n}{\partial n_s}-\psi_n\frac{\partial\psi_\lambda}{\partial n_s}\right)dS. \quad . \quad (9)$$

Here n_s denotes the outward normal at the particle surface into the region of integration, which is the inward normal to the surface.

Suppose that the dielectric constant of the colloidal particle is D^I and ψ^I represents the potential function inside the particle*. Since there is no charge inside the particle

$$\nabla^2\psi^I=0, \qquad \nabla^2\psi^I_n=0, \qquad \nabla^2\psi^I_\lambda=0, \quad . \quad . \quad . \quad . \quad (10)$$

and on the surface we have the boundary conditions

$$\psi=\psi^I, \qquad D\frac{\partial\psi}{\partial n_s}-D^I\frac{\partial\psi^I}{\partial n_s}=-4\pi\nu\lambda e_1. \quad . \quad . \quad . \quad . \quad (11)$$

We proceed to evaluate the various contributions to (6), which come from the surface integrals in (9). Interchanging the order of integration with respect to λ and S and then integrating by parts,

$$-\frac{D}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi_\lambda\frac{\partial\psi_n}{\partial n_s}dS=-\frac{D}{4\pi}\int_S\left[\psi\frac{\partial\psi_n}{\partial n_s}\right]_{\lambda=\lambda'}dS+\frac{D}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi\frac{\partial\psi_{n\lambda}}{\partial n_s}dS \quad . \quad . \quad . \quad (12)$$

It is seen from (11) that $\partial\psi/\partial n_s=0$ when $\lambda=0$, although it is possible that $\psi\neq0$. On applying the boundary conditions (11), and noting that $\nu_\lambda=0$, the right-hand side of (12) can be expressed as

$$\lambda' e_1\int_S\psi_{\lambda=\lambda'}\nu_n\,dS-\frac{D^I}{4\pi}\int_S\left[\psi^I\frac{\partial\psi^I_n}{\partial n_s}\right]_{\lambda=\lambda'}dS-e_1\int_0^{\lambda'}d\lambda\int_S\psi\nu_n\,dS$$
$$+\frac{D^I}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi^I\frac{\partial\psi^I_{n\lambda}}{\partial n_s}dS, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12i)$$

where $\psi_{\lambda=\lambda'}$ is the (surface) potential at stage λ' and position **s**. This disposes of the first term in the surface integral in (9). The second term in this surface integral yields the contribution

$$+\frac{D}{4\pi}\int_0^{\lambda'}\partial\lambda\int_S\psi_n\frac{\partial}{\partial\lambda}\left(\frac{\partial\psi}{\partial n_s}\right)dS=-e_1\int_0^{\lambda'}d\lambda\int_S\psi_n\nu\,dS+\frac{D^I}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi^I_n\frac{\partial\psi^I_\lambda}{\partial n_s}dS \quad . \quad . \quad . \quad (13)$$

to (*b*). If we now combine the three integrals in (12*i*) and (13) which involve the function ψ^I and make use of the first of the boundary conditions (11), then a simple integration by parts yields the expression

$$-\frac{D^I}{4\pi}\int_0^{\lambda'}d\lambda\int_S\left(\psi^I_\lambda\frac{\partial\psi^I_n}{\partial n_s}-\psi^I_n\frac{\partial\psi^I_\lambda}{\partial n_s}\right)dS=0. \quad . \quad . \quad . \quad . \quad (14)$$

---

* In the case of a metallic particle, we put ψ^I=constant.

This follows on application of Green's theorem and of the equation (10). Also, we have two terms, one from ($12i$) and the other from (13), which cancel out with the contribution from the surface integral on the right of (7). We finally derive that

$$\frac{\partial \mathrm{F}_e(\lambda', n)}{\partial n} = \lambda' e_1 \int_{\mathrm{S}} \psi_{\lambda=\lambda'} \nu_n \, d\mathrm{S}. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The formula (15) can be given a simple interpretation. When we bring up δn ions, each of charge $\lambda' e_1$, to the surface, the increment in the density of ions at position $\mathbf{s}$ is $\delta\nu = \nu_n \delta n = \nu_0(\lambda', \mathbf{s})\, \delta n$. The contribution to the change in electrical energy from the surface element $d\mathrm{S}$ is $\lambda' e_1 \psi_{\lambda=\lambda'} \delta\nu \, d\mathrm{S}$ which is the electrical work done in bringing $\delta\nu$ ions to the element $d\mathrm{S}$. In particular, if the density of ions is constant over the surface, then $\nu_0(\lambda', \mathbf{s}) = 1/\mathrm{S}$ and (15) becomes

$$\frac{\partial \mathrm{F}_e(\lambda, n)}{\partial n} = \lambda' e_1 \psi_{\mathrm{ave}}(\lambda', n), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where $\psi_{\mathrm{ave}}(\lambda', n)$ is the average potential on the surface. If we define the average potential in terms of the weight function $\nu_n = \nu_0(\lambda', \mathbf{s})$, then (16) can be interpreted as covering the general case (15). We now integrate with respect to n from 0 to n', say, and so obtain

$$\mathrm{F}_e(\lambda', n') = \lambda' e_1 \int_0^{n'} \psi_{\mathrm{ave}}(\lambda', n) \, dn, \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where, in general,

$$\psi_{\mathrm{ave}}(\lambda', n) = \int_{\mathrm{S}} \psi_{\lambda=\lambda'} \nu_0(\lambda', \mathbf{s}) \, d\mathrm{S}. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The relation (17), which was already implied in a paper by Derjaguin (1940), was proposed without proof by the author in an earlier paper (Levine 1948). The right-hand side of (17) is the usual formula from electrostatics for the work done in bringing n' ions, each of charge $\lambda' e_1$, by infinitesimal amounts from the solution to the surface of the particle, the potential of which is $\psi_{\mathrm{ave}}(\lambda', n)$.

§ 3. ENERGY OF TWO PARTICLES.

We can readily extend the result (17) to a dilute sol containing P particles. Suppose that there are n'_p ions of type 1 on the surface of the pth particle. We imagine that these ions are transported from the solution to the surface at the same rate for all the particles and that the fraction ξ, which replaces the variable n, denotes the stage reached in this process. Thus $n_p = \xi n'_p$ will be the number of ions on particle p at stage ξ. The potential ψ_p and the surface density of ions ν_p on particle p will be considered as functions of λ, ξ and the position on the particle surface. The derivation of a simplified expression for the free energy follows the same steps as for a single particle provided we replace e_1 by $n'_p e_1$ for the pth particle, differentiate with respect to ξ instead of n and sum over all the particles. The electrical free energy of the P double layers at stage λ' can be finally expressed in the form

$$\lambda' e_1 \sum_{p=1}^{\mathrm{P}} n_p \int_0^1 \psi_p^{\mathrm{ave}}(\lambda', \xi n_p')\, d\xi = \lambda' e_1 \sum_{p=1}^{\mathrm{P}} \int_0^{n_p'} \psi_p^{\mathrm{ave}}(\lambda', n_p)\, dn_p, \quad . \quad (19)$$

where $\psi_p^{\mathrm{ave}}(\lambda', n_p)$ is the average potential on particle p.

In particular, we can apply this formula to two parallel plates, each of area A, at separation R. Since we are dealing with closed surfaces it is necessary to consider both sides of each plate and also the edges. The latter, however, can be neglected for sufficiently thin plates. In addition, we need not concern ourselves with the contribution to the free energy from the outer sides, since this will not vary with the separation R. Making use of equation (9) in a previous paper (Levine 1948) the electrical free energy due to the excess charge in the two inner double layers can be written as

$$\mathrm{F}_e(\lambda', \nu', \mathrm{R}) = 2\mathrm{A}\lambda' e_1 \int_0^{\nu'} \Phi(\lambda', \nu, \mathrm{R})\, d\nu, \quad . \quad . \quad . \quad . \quad (20)$$

where $\lambda' e_1 \Phi(\lambda', \nu, \mathrm{R})$ is the potential of the mean force acting on an adsorbed ion, due to the excess ions in the layers; the zero of this potential is taken in the interior of the solution. For finite plates, both ν and $\Phi(\lambda', \nu, \mathrm{R})$ would not be constant over the inner faces because of the edge effects and we would have to introduce averages. However, this correction is neglected here. The result (20) follows from the general theory based on the methods of statistical mechanics, which is developed elsewhere (Levine 1949b). If we now introduce the approximation of the Poisson-Boltzmann equation and identify $\Phi(\lambda', \nu, \mathrm{R})$ with the potential function $\psi_p(\lambda', n_p)$, $(p=1, 2)$, then the formula (20) is immediately derived from (19). In particular, if ν is sufficiently small, $\Phi(\lambda', \nu, \mathrm{R})$ can be assumed proportional to ν, in which case we have

$$\mathrm{F}_e(\lambda', \nu', \mathrm{R}) = \mathrm{A}\lambda' e_1 \nu' \Phi(\lambda', \nu', \mathrm{R}). \quad . \quad . \quad . \quad . \quad (20i)$$

This result was already suggested in an earlier paper (Levine 1948).

When the linear Debye-Huckel equation* is applied to two (infinite) parallel plates, the potential between the plates is given by

$$\psi(\lambda, x) = \frac{4\pi\nu e_1}{\mathrm{D}\kappa_0} \frac{\cosh \lambda\kappa_0 x}{\sinh \lambda\kappa_0 \mathrm{R}/2}, \qquad \kappa_0^2 = \frac{4\pi}{\mathrm{D}k\mathrm{T}} \sum_{i=1}^{s} n_i^0 e_i^2, \quad . \quad . \quad (21)$$

where x is measured from the median plane, so that we assume $\psi(\lambda, \mathrm{R}/2) = \Phi(\lambda, \nu, \mathrm{R})$. We find that an additional term appears on the right-hand side of (20i), namely $-8\pi\mathrm{A}\nu'^2 e_1^2/\mathrm{D}\kappa_0^2\mathrm{R}$. This difficulty was met in a previous paper (Levine 1946) and a lengthy discussion by Overbeek (1946) and the author (Levine 1946a) followed concerning this discrepancy. Although the essential basis for the appearance of this extraneous term was first proposed by Overbeek, no conclusive proof that (20*i*) is actually the correct result was given. We are now in a position to show clearly the origin of this additional term.

---

* We expand $f(\lambda\psi)$ in powers of ψ and retain the first term only, namely $f(\lambda\psi) \approx \lambda\kappa_0^2\psi$.

When all the ions are discharged, the formula (21) yields the condition

$$\lim_{\lambda=0} \lambda\psi(\lambda, x)=8\pi\nu e_1/D\kappa_0^2R=m, \quad \text{say}, \qquad (22)$$

throughout the region between the two plates and the fact that the potential $\psi(\lambda, x)\to\infty$ as $1/\lambda$ when $\lambda\to 0$ creates a difficulty. Use of the Boltzmann distribution law yields the expression $n_i^0 \exp(-e_i m/kT)$ for the limiting density (at $\lambda=0$) of i ions at position x. But if the short-range van der Waals forces near the particle walls are neglected, then necessarily the completed discharged ions should be uniformly distributed throughout the dispersion medium, which means that the condition $\lim_{\lambda=0} \lambda\psi(\lambda, x)=0$ must be satisfied, in contradiction to (22).

The analysis for a single plate developed in the previous section can be applied to each plate, provided the volume of the diffuse outer layer does not extend to infinity but to the median plane. If now $\lim_{\lambda\to 0} \lambda\psi\neq 0$ then the integral in (5) still converges but this is not true of each of the separate terms obtained by substituting (4) into (5). To overcome this difficulty of divergent integrals, we replace the lower limit of integration $\lambda=0$ by $\lambda=\epsilon$, where ϵ is a small positive number, and after integration let $\epsilon\to 0$. An additional term is obtained on the right of (12), namely*

$$-\frac{D}{4\pi}\lim_{\lambda\to 0}\int_s \psi\frac{\partial\psi_n}{\partial n_s}\,dS=-\lim_{\lambda\to 0}\lambda e_1\psi_{\text{ave}}(\lambda, n), \qquad (23)$$

which does not vanish, since $\partial\psi/\partial n_s$ behaves as λ whereas ψ behaves as $1/\lambda$ when $\lambda\to 0$. The form on the right of (23) is obtained when we apply the condition that $F_e(\lambda, n)\to 0$ as $\lambda\to 0$. We now need to introduce a corresponding correction to the right-hand side of our basic relation (17), which becomes

$$F_e(\lambda', n')=\lambda' e_1\int_0^{n'}\psi_{\text{ave}}(\lambda', n)\,dn-\lim_{\lambda\to 0}\lambda e_1\int_0^{n'}\psi_{\text{ave}}(\lambda, n)\,dn. \qquad (24)$$

The second term on the right of (24) was suggested by Overbeek (1946) although his explanation was not quite adequate. In the case of each of our two parallel plates, this new term becomes

$$-\frac{DA}{4\pi}\lim_{\lambda\to 0}\int_0^{\nu'}\psi\left(\lambda, \frac{R}{2}\right)\left(\frac{\partial\psi_\nu(\lambda, x)}{\partial x}\right)_{x=R/2}d\nu$$
$$=-\lim_{\lambda\to 0}A\lambda e_1\int_0^{\nu'}\psi\left(\lambda, \frac{R}{2}\right)d\nu=-\frac{4\pi A\nu'^2 e_1^2}{D\kappa_0^2 R}, \qquad (25)$$

---

* We note that at the median plane, $\partial\psi/\partial n_s=0$, since the potential is a minimum.

which is precisely half the additional (extraneous) term appearing on the right of (20i). The anomaly arises because we are considering two plates of infinite extent. This model cannot be used to describe two finite plates in the neighbourhood of $\lambda=0$ since it prevents the diffusion of the excess charge of the diffuse layers between the plates into the bulk of the solution as $\lambda\to 0$. The necessary condition $\lim_{\lambda\to 0}\lambda\psi=0$ ensures that the expression (25) always vanishes and hence our formula (17) applies to each of the finite plates.

For two identical particles, symmetrically orientated with respect to the line joining their centres of mass, we shall denote the (weighted) average potential over each particle, which is defined in (18), by $\psi_{\text{ave}}(\lambda', n, \text{R})$. It follows from (17) that the electrical free energy of the double layers of the two particles is given by

$$\text{F}_e(\lambda', n', \text{R})=2\lambda' e_1\int_0^{n'}\psi_{\text{ave}}(\lambda', n, \text{R})\,dn. \quad . \quad . \quad . \quad . \quad (26)$$

This is the simplified form that we set out to derive in this section.

§4. Casimir's Theorem of Equivalence.

This theorem of the equivalence of the two methods, employed by Verwey and Overbeek (1948) to evaluate the free energy of the double layer, can be derived from our general results in §2. It will be sufficient to consider a single particle, since the argument is readily extended to an arbitrary number of particles. Also we shall first derive the Casimir theorem in the special circumstances when (16) is replaced by

$$\frac{\partial \text{F}_e(\lambda', n)}{\partial n}=\lambda' e_1\psi(\lambda', n), \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where $\psi(\lambda', n)$ denotes the constant potential at the surface, *i. e.* it does not depend on the position $\mathbf{s}$, but it is a function of λ' and n. It should be noted that n is treated as an arbitrary parameter in (27). The following considerations show that (27) is valid for any n in two particular cases.

Suppose that we replace the general form (1) for ν by

$$\nu=n^*(\lambda')\nu_0^*(\lambda', \mathbf{s}), \qquad \int_{\text{S}}\nu_0^*(\lambda', \mathbf{s})\,d\text{S}=1, \quad . \quad . \quad . \quad . \quad (1i)$$

such that in (18), $\psi_{\lambda=\lambda'}$ is uniform over the particle surface. In these circumstances, it might appear that (27) applies, but we find that in general, this is not the case. For let the arbitrary function $\nu_0(\lambda', \mathbf{s})$ in (18) be replaced by $\nu_0^*(\lambda', \mathbf{s})$ but at the same time n be given any value. Then the surface potential $\psi_{\lambda=\lambda'}$ will depend on n and, in general, also on the position $\mathbf{s}$. If we now introduce $n=n^*(\lambda')$ as well as the distribution $\nu_0^*(\lambda', \mathbf{s})$, then $\psi_{\lambda=\lambda'}$ does become uniform over the surface. It follows that for this particular value of n (and distribution $\nu_0^*(\lambda', \mathbf{s})$), $\psi_{\text{ave}}(\lambda', n^*(\lambda'))$ can be identified with $\psi(\lambda', n^*(\lambda'))$; in other words the relation (27) becomes valid if we susbstitute $n=n^*(\lambda')$. The problem is whether $\psi_{\lambda=\lambda'}$ is constant over the surface when the relative distribution is again $\nu_0^*(\lambda', \mathbf{s})$ but n is arbitrary. One case where this is true is that of small n, when (2) reduces to the linear Debye-Huckel equation, with solution $\psi = n\phi$, where the function ϕ depends on the form of $\nu_0^*(\lambda', \mathbf{s})$ but not on n. For we can choose the distribution $\nu_0^*(\lambda', \mathbf{s})$ such that the value of ϕ at the surface (at stage λ') is constant, independent of the position $\mathbf{s}$. Denoting this constant value by $\phi_{\lambda=\lambda'}$, (27) will apply, with $\psi(\lambda', n) = n\phi_{\lambda=\lambda'}$. However, if n is not small, the essential property that ψ be proportional to n no longer holds and (27) is not valid for a particle of general shape. The other case arises when we are dealing with a spherical, cylindrical or plate-like particle, or with two parallel plates, provided the surface density of ions is uniform (equal to n/S) and any edge effects due to the finite size of the particles are neglected.

Now in §2, we can treat λ' as a parameter, just like n, and we obtain the relations

$$\nabla^2\psi_{\lambda'} = \lambda^2 f'\psi_{\lambda'}, \qquad \nabla^2\psi^{\mathrm{I}}_{\lambda'} = 0,$$

where the subscript λ' denotes partial differentiation. Using precisely the same method as in §2, we can derive quite generally

$$2\int_0^{\lambda'} \frac{1}{\lambda} \frac{\partial \mathrm{E}(\lambda, n\,;\ \lambda')}{\partial \lambda'}\, d\lambda = \lambda' e_1 \int_{\mathrm{S}} \psi_{\lambda=\lambda'} \nu_{\lambda'}\, d\mathrm{S} = \lambda' e_1 \Psi(\lambda', n), \text{ say.} \quad (28)$$

It is observed that (28) immediately follows from (6) and (15) when n is replaced by λ'. When (27) applies, then $\psi(\lambda', n) = \psi_{\lambda=\lambda'}$ and hence

$$\Psi(\lambda', n) = \psi(\lambda', n) \frac{\partial}{\partial \lambda'}\left[\int_{\mathrm{S}} \nu\, d\mathrm{S}\right] = 0, \quad . \quad . \quad . \quad . \quad (28i)$$

since the surface integral in (28*i*) is equal to n.

Since it is simpler to treat the case where the potential at the surface is constant (equal to $\psi(\lambda', n)$), we begin with the relation

$$2\int_0^{\lambda} \frac{1}{\lambda} \mathrm{E}(\lambda, n'\,;\ \lambda')\, d\lambda = \lambda' e_1 \int_0^{n'} \psi(\lambda', n)\, dn, \quad . \quad . \quad . \quad (29)$$

which follows from (5) and (27). Remembering that λ' and n' are considered to be independent of one another, we differentiate both sides of (29) partially with respect to λ', make use of the relations (28) and (28*i*), and then put $\lambda' = \lambda$. In this way we obtain

$$\frac{2}{\lambda} \mathrm{E}(\lambda, n'\,;\ \lambda) = e_1 \int_0^{n'} \psi(\lambda, n)\, dn + \lambda e_1 \int_0^{n'} \psi_\lambda(\lambda, n)\, dn. \quad . \quad . \quad (30)$$

This is an identity in λ and n', and in particular, will be true if n' is chosen that function of λ which is defined by requiring that $\psi(\lambda, n') = \psi_0$, a constant. This function is given by the inverse relation which we write as

$$n' = n(\lambda, \psi_0) \, . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

If the number of surface ions is given by (31), the electrostatic energy of the double layer can be expressed as

$$E(\lambda, n(\lambda, \psi_0);\ \lambda)=\tfrac{1}{2}\lambda e_1\psi_0 n(\lambda, \psi_0)+\tfrac{1}{2}\int_V [\psi\rho]_{\psi_0}\, dv, \quad . \quad . \quad . \quad (32)$$

where the surface potential is constant (equal to ψ_0) when integrating over the volume V on the right-hand side. If we also substitute (31) for n' into (30), then it follows that

$$\frac{1}{\lambda}\int_V [\psi\rho]_{\psi_0} dv=-e_1\int_0^{\psi_0} n(\lambda, \psi)\, d\psi+\lambda e_1\int_0^{n(\lambda, \psi_0)} \psi_\lambda(\lambda, n)\, dn. \qquad (33)$$

The first term on the right of (33) has been written in place of

$$e_1\int_0^{n(\lambda, \psi_0)} \psi(\lambda, n)\, dn-e_1\psi_0 n(\lambda, \psi_0).$$

Since $\psi_0=\psi(\lambda, n(\lambda, \psi_0))$, we have the relations

$$\int_0^{n(\lambda, \psi_0)} \psi_\lambda(\lambda, n)\, dn=\frac{\partial}{\partial\lambda}\left[\int_0^{n(\lambda, \psi_0)} \psi(\lambda, n)\, dn-\psi_0 n(\lambda, \psi_0)\right]=-\frac{\partial}{\partial\lambda}\int_0^{\psi_0} n(\lambda, \psi)\, d\psi,$$

and hence (33) becomes

$$\frac{1}{\lambda}\int_V [\psi\rho]_{\psi_0} dv=-\frac{\partial}{\partial\lambda}\left[\lambda e_1\int_0^{\psi_0} n(\lambda, \psi)\, d\psi\right]. \quad . \quad . \quad . \quad (34)$$

Integrating with respect to λ from 0 to λ', we finally derive

$$\int_0^{\lambda'} \frac{d\lambda}{\lambda}\int_V [\psi\rho]_{\psi_0} dv=-\lambda' e_1\int_0^{\psi_0} n(\lambda', \psi)\, d\psi+\lim_{\lambda\to 0} \lambda e_1\int_0^{\psi_0} n(\lambda, \psi)\, d\psi. \quad (35)$$

For plate-like particles, the second term on the right of (35) vanishes and we obtain the formula derived by Casimir (1948). For a cylindrical particle, the solution of the linear Debye-Huckel equation yields the formulæ

$$\psi(\lambda, n)=\frac{2ne_1}{D\kappa_0 a}\,\frac{K_0(\lambda\kappa_0 a)}{K_1(\lambda\kappa_0 a)}, \qquad n(\lambda, \psi_0)=\frac{D\kappa_0 a\psi_0}{2e_1}\,\frac{K_1(\lambda\kappa_0 a)}{K_0(\lambda\kappa_0 a)}, \quad . \quad (36)$$

where n is the number of surface ions per unit length of particle, a is the radius of the circular cross-section and K_0 and K_1 are Bessel functions of purely imaginary argument. Since $K_0(\lambda\kappa_0 a)$ and $K_1(\lambda\kappa_0 a)$ behave as $\ln\lambda$ and $1/\lambda\kappa_0 a$ respectively when $\lambda\to 0$, the second term on the right of (35) again vanishes, although $n\to\infty$ (which is, of course, physically impossible).

In the case of a spherical particle of radius a, use of the linear Debye-Huckel equation for ψ yields the formula

$$\psi=\psi_0 a \exp\left[-\lambda\kappa_0(r-a)\right]/r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

for the potential distribution in the double layer, where r is the distance from the centre of the sphere. Thus

$$\psi(\lambda, n)=\frac{n\lambda e}{Da(1+\lambda\kappa_0 a)}, \qquad n(\lambda, \psi_0)=\frac{Da\psi_0}{\lambda e_1}(1+\lambda\kappa_0 a), \quad . \quad . \quad (38)$$

so that $n(\lambda, \psi_0)\to\infty$ as $\lambda\to 0$. We now find that

$$\lim_{\lambda\to 0} \lambda e_1 \int_0^{\varphi_0} n(\lambda, \psi)\, d\psi = \mathrm{D}a\psi_0^2/2. \quad . \quad . \quad . \quad . \quad . \quad (39)$$

If we substitute (37) and the corresponding volume density of charge

$$\rho = -(\mathrm{D}/4\pi)\nabla^2\psi = -\mathrm{D}\lambda^2\kappa_0^2\psi/4\pi, \quad . \quad . \quad . \quad . \quad . \quad (40)$$

which follows from (2) and the Debye-Huckel approximation $f(\lambda\psi)=\lambda\kappa_0^2\psi$, the left-hand side of (35) becomes

$$-\mathrm{D}a^2\psi_0^2 \int_0^{\lambda'} \lambda \exp(2\lambda\kappa_0 a)\, d\lambda \int_a^\infty \exp(-2\lambda\kappa_0 r)\, dr = -(\mathrm{D}a\psi_0^2/2)\lambda'\kappa_0 a. \quad (41)$$

However, it follows from (38) that the first term on the right of (35) is

$$-(\mathrm{D}a\psi_0^2/2)\,(1+\lambda'\kappa_0 a)\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

No mention of this exception to the Casimir theorem is made in the book of Verwey and Overbeek. The omission by Casimir (1948) of the second term on the right of (35) is made at the end of his proof when he assumes (in his notation) $\partial\chi(\lambda=0)/\partial n=0$, *i.e.* the derivative of the potential in the direction normal to the particle surface vanishes at $\lambda=0$ *. However, this condition is not satisfied for a spherical particle since it follows immediately from (37) that $\lim\limits_{\lambda\to 0} (\partial\psi/\partial r)_{r=a} = -\psi_0/a$. This derivative of the potential does vanish at the surface of a plate or cylinder.

It is possible to obtain a generalization of the relation (35) for a particle of any shape, the surface of which need not be at a constant potential. On the right-hand side of (29) we replace the function $\psi(\lambda', n)$ by $\psi_{\text{ave}}(\lambda', n)$ which is defined by (18), with the distribution $\nu_0(\lambda', \mathbf{s})$ chosen arbitrarily. Differentiating the modified form of (29) with respect to λ', we again put $\lambda'=\lambda$. It is now assumed that n' is chosen, such a function of λ that the quantity $\psi_{\text{ave}}(\lambda, n')$ becomes equal to a constant ψ_0, independent of λ. This function will be denoted as

$$n' = n_{\text{ave}}(\lambda, \psi_0), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31\,i)$$

which replaces (31). (Since $\psi_{\text{ave}}(\lambda, n')$ refers to stage λ, it is defined by the weight function $\nu_0(\lambda, \mathbf{s})$ in (18).) The operations performed on the right-hand side of (29) are precisely the same as those in the analysis above. Turning our attention to the left-hand side of (29), it is seen that we are charging the ions from 0 to λ', subject to a fixed surface distribution of ions, given by $\nu = n'\nu_0(\lambda', \mathbf{s})$. Also, an additional term now appears on the left of (30), namely $\lambda e_1 \Psi(\lambda, n')$, which will not vanish in the general case. When we reach the equation which is the generalization of (35), namely

$$\int_0^{\lambda'} \frac{d\lambda}{\lambda} \int_{\mathrm{V}} [\psi\rho]_{\varphi_{\text{ave}}=\varphi_0}\, dv = -e_1 \int_0^{\lambda'} \Psi(\lambda, n_{\text{ave}}(\lambda, \psi_0))\, d\lambda - \lambda' e_1 \int_0^{\varphi_0} n_{\text{ave}}(\lambda, \psi)\, d\psi$$

$$+\lim_{\lambda\to 0} \lambda e_1 \int_0^{\lambda_0} n_{\text{ave}}(\lambda, \psi)\, d\psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

---

* Subject to fixed surface potential (4_0).

the integration with respect to λ on the left-hand side is being carried out under the condition that, at each stage λ, the average potential over the particle surface is $\psi_{\text{ave}}(\lambda, n_{\text{ave}}(\lambda, \psi_0))=\psi_0$, a constant. This result reduces to the Casimir theorem for a particle of general shape in the following special circumstances. The distribution $\nu_0(\lambda, \mathbf{s})$ (at stage λ) and the total number of surface ions n are chosen in such a way that the surface potential is constant over the surface and equal to ψ_0. We can denote this distribution by $\nu_0^*(\lambda, \mathbf{s})$ and the value of n by $n^*(\lambda, \psi_0)$, which now replaces $n_{\text{ave}}(\lambda, \psi_0)$. Then it follows from (28) that $\Psi(\lambda, n^*(\lambda, \psi_0))=0$ and the relation (18) yields (at $\lambda'=\lambda$) $\psi_{\text{ave}}(\lambda, n^*(\lambda, \psi_0))=\psi_{\lambda=\lambda}=\psi_0$. Hence (43) takes on the same form as (35), which is the required result.

§ 5. DISCUSSION.

We can suggest the following physical interpretation of our theorem which is expressed in equations (5), (17) and (18). All methods of picturing the formation of the electric double layers of the particles should lead to the same value for the free energy provided such methods describe thermodynamically reversible processes. Suppose now that the surface density of ions is treated as a thermodynamic variable, which we can change at will. Then the charging process introduced in the definition (5) of the free energy is carried out under equilibrium (reversible) conditions and is therefore valid. Our second method which leads to the expression (17) for the free energy is also valid provided the electric work done in bringing the ions to the surface is against the potential of the mean force acting on the ions. When we use the approximation of the Poisson-Boltzmann equation (the potential of the mean force being identified with the mean potential) and also restrict ourselves to very dilute sols, then this additional condition in the second method is apparently accounted for.

It might appear that we have introduced some restriction by keeping the distribution (and hence number) of surface ions fixed when charging the double layers. However, this is not the case and the explanation depends on two theorems in (classical) statistical mechanics. Firstly, it has been shown by Onsager (1933) that it is permissible to apply the charging process to a sub-space of the positional part of the phase space of the ions. Such a sub-space is defined by a fixed (but arbitrary) distribution of surface ions but *not*, in general, by choosing the surface potential as constant. Secondly, the free energy has an extremely sharp minimum at the equilibrium distribution of surface ions. This latter property implies that the size of the particles is so large that we are dealing with a macrospocic system. By treating a large number of particles we have, in effect, macroscopic particle surfaces. (We ignore here the complication which arises when the adsorbed ions form part of the crystal lattice of the particles and which is briefly discussed elsewhere (Levine 1949 a)).

The two methods of Verwey and Overbeek for evaluating the free energy lead to equation (35), with the second term on the right-hand side absent. In developing a physical argument to interpret the equivalence of their methods, these authors introduce the apparently arbitrary assumption that the charging process be carried out under the condition of constant surface potential. Previously (Levine 1946 b) the author had raised the question as to whether the equilibrium conditions were being maintained in this charging method. In the case of spherical particles, for example, the number of surface ions n becomes infinite as $\lambda \to 0$, when the surface potential is kept fixed. As a consequence, the two methods originally proposed by Verwey and Overbeek yield different results and the Casimir theorem of equivalence has to be modified by introducing the second term on the right of (35).

It appears, therefore, that equilibrium conditions are not kept in the process of charging at constant surface potential. In the opinion of the author, little or no physical significnace can be attached to the artificial process visualized by Verwey and Overbeek. Nevertheless it is possible to obtain the correct expression for the free energy by this method, although not in all cases (*e. g.* spheres). The reason for the success of this approach is to be found in the nature of the theorems relating to the free energy, which are based on the use of the Poisson-Boltzmann equation and the restriction to very dilute sols. A number of these theorems have been developed in this paper, and others will be given later. Thus, we regard the equivalence theorem of Casimir as a convenient mathematical relation.

In conclusion, the author wishes to thank Professor J. Th. G. Overbeek for helpful correspondence on a number of aspects in this paper.

REFERENCES.

Casimir, H. G. B., 1948, in *Theory of Stability of Lyophobic Colloids, Elsevier* p. 63, by Verwey and Overbeek.

Derjaguin, B., 1940, *Trans. Faraday Soc.*, **36**, 203.

Grimley, T. B., and Mott, N. F., 1947, *Faraday Soc. Discussion*, **1**, 3.

Levine, S., 1946 a, *Trans. Faraday Soc., Swelling and Shrinking issue*, **42B**, 102; 1946 b, *Ibid.*, 128; 1948, *Trans. Faraday Soc.*, **44**, 833; 1949 a, *J. Colloid Science* (in press); 1949 b, submitted to *Proc. Camb. Phil. Soc.*

Onsager, L., 1933, *Chem. Rev.*, **13**, 73.

Overbeek, J. Th. G., 1946, *Trans. Faraday Soc Swelling and Shrinking issue*, **42B**, 126.

Verwey, E. J. W. and Overbeek, J. Th. G., 1948, *Theory of the Stability of Lyophobic Colloids, Elsevier.*

V. *Noise in Ionization Chamber Pulse Amplifiers.*

By R. Wilson, D.Phil.,
Clarendon Laboratory, Oxford*.

[Received September 15, 1949.]

Abstract.

The general problem of the detection of heavy ionizing particles, in the presence of γ radiation and background noise from the amplifier, is considered. It is shown that the maximum possible signal to noise ratio is obtained if the input pulse is shaped to a form $\exp\{-|t-T_D|/\tau\}$. The signal to noise ratio is also calculated for other common pulse shapes.

A practical amplifier circuit is discussed which may be used to detect 100 kv. energy protons in the presence of intense γ radiation.

Introduction.

The signal from an ionization chamber is a pulse of current which flows in an external circuit as the movement of ions induces charge on the electrodes. If free electrons are collected, and the counter geometry is properly designed, this time may be as low as $0{\cdot}1\,\mu$s. which is less than the usual amplifier time constants. The total charge which passes is proportional to the energy of the initial ionizing particle.

It is desired to amplify this current pulse and to distinguish it from background " noise ". This background noise is due to the superposition of many smaller pulses ; some due to electrons in the counter, others due to the spontaneous thermal agitation of electrons in the electrical circuit (Johnson noise) or the spontaneous movement of electrons in the amplifying valves (shot effect).

The wanted current pulse, like all other current pulses, is transformed into a step pulse of potential on the grid of the first valve, by charging the stray capacity to earth of the ionization chamber and valve. The grid leak of this valve discharges the condenser after a long time, but the grid leak is usually made large in order not to produce appreciable resistor noise, so that this time is sufficiently long to be regarded as infinite.

The amplifier will distort this pulse of potential. Part of this distortion may be deliberate, part caused by defects in the amplifier. It is convenient, however, to consider an ideal, non-distorting, linear amplifier and certain linear pulse-shaping networks which distort this and every other pulse. Then the problem is to decide upon the best way to distort this pulse in order to discriminate against the background noise. It is convenient

---

* Communicated by the Author.

to "refer" the various sources of background noise to the input as an "equivalent noise charge". This is a fictitious noise charge on the ionization chamber electrodes, which produces the same signal at the amplifier output as the noise.

It will be shown that with a conventional resistance—capacitance pulse shaping network the best signal to noise ratio is obtained when the time constants of "differentiation" and "integration" are equal; that an improvement of 20 per cent is possible by using a shorted delay time for shaping, and that a further improvement of 15 per cent is possible by using a network to shape the pulse to a form $\exp\{-|t-T_D|/\tau\}$. The overall improvement of 35 per cent may be very important.

ANALYTICAL EXPRESSION FOR SIGNAL TO NOISE RATIO.

It will here be assumed that the duration of the current pulse is short compared with the pulse shaping time constants so that the input may be regarded as a function $Q\delta(t)$ of time—where $\delta(t)$ is Dirac's δ function defined by $\delta(t)=0$ for $t\neq 0$ and $\int_{-\infty}^{+\infty}\delta(t)\,dt=1$.

If this assumption is not true (as for example with ion collection in counters) the pulses from the ionization chamber will be attenuated more than the background pulses which is undesirable if a high signal to noise ratio is wanted.

The potential charge on the first valve grid is then $\frac{Q}{c}H(t)$ where $H(t)$ is Heaviside's unit function.

$$H(t)=1 \qquad t>0,$$

$$H(t)=0 \qquad t<0,$$

$$\text{and} \quad \frac{d\{H(t)\}}{dt}=\delta(t).$$

All pulses will be modified by the shaping network. If the input potential pulse is $H(t)$, then the output pulse is a function $f(t)$ of the shaping network. If the input potential pulse is $\delta(t)$, the output pulse will be $d/dt\{f(t)\}=f'(t)$.

The amplifier noise is due to the superposition of the effects of individual electrons in those amplifier elements which are at a low signal level—the input grid leak, the first valve and the first valve anode load. When the ionization chamber is used to count heavy ionizing particles in the presence of intense γ radiation the secondary electron pulses in the chamber also cause a background.

The noise is often expressed in terms of the amplifier bandwidth, here this practice is reversed, and the noise is expressed in terms of the output pulse shape, which is considered to be more fundamental in a pulse amplifier than the bandwidth. This involves a more fundamental consideration of the noise problem.

The valve shot noise may be directly expressed as the superposition of pulses caused by individual electrons in the current stream. Each electron will produce a current pulse as it reaches the electrode; since the transit time ($\sim 10^{-9}$ secs.) is small compared with the amplifier time constants, this may be regarded as a pulse of $e\delta(t)$.

Here can be distinguished two classes of noise. The valve anode has a resistive load R, so that a potential pulse $\mathrm{R}e\delta(t)$ is developed across the load. The superposition of these pulses will be called class I noise. On the other hand, the valve grid has a capacitative load C, and a potential pulse $e\mathrm{H}(t)/\mathrm{C}$ will be developed across the load. The superposition of these other pulses will be called class II. noise.

In practice it is desired to reduce the signal level at which accurate counting is possible. An arbitrary maximum rate of counting of noise peaks is set for each experiment, and the best amplifier conditions must be found in order to reduce the signal level at which this counting rate occurs. A useful concept for this problem is the root mean square noise fluctuation. If the noise current is $\mathrm{I}(t)$ the r.m.s. noise fluctuation σ is given by $\sqrt{\overline{(\mathrm{I}(t)-\overline{\mathrm{I}(t)})^2}}$.

If the rate of arrival of noise pulses is large compared with the frequencies passed by the amplifier, the rate of counting of noise peaks may be expressed in terms of σ, the signal level V, and the amplifier bandwidth. (See equation (20) later). The problem of minimizing the signal level for a constant rate of counting is difficult, but the amplifier bandwidth only enters into the counting rate in a linear term, whereas σ enters in an exponential term. A close approximation, increasingly accurate as V/σ increases, will be to minimize σ. A rough check shows that the pulses should be about 10 per cent wider than given by this criterion, but this would alter the minimum detectable signal by only 1 per cent.

The assumption that the rate of arrival of noise pulses will be greater than the frequencies passed by the amplifier will usually hold for shot noise from both the anode and the grid, and for thermal (resistor) noise. But the γ ray background may not be sufficiently intense for the pulses to "pile up" in this way. In this case the γ rays will be counted as separate pulses of height ke.

Change of pulse shape will not then alter the rate of counting of background at all; so that the γ ray background must be treated as distinct from all other forms of noise. If the pulses are small, their effect will be inappreciable, and if large, they will be similar to unwanted heavy ionizing particles.

An intermediate case will often arise where a certain amount of pile-up occurs which is difficult to treat theoretically, especially since the other forms of noise are added.

If these limitations do not apply, we may proceed by finding the root mean square fluctuation from Campbell's Theorem (Campbell 1909).

Campbell's theorem states that if a noise current $\mathrm{I}(t)$ is caused by the superposition of ν events per second, each causing an effect $f(t)$ then the root mean square fluctuation σ is given by

$$\sigma^2=\overline{(\mathrm{I}(t)-\overline{\mathrm{I}(t)})^2}$$

$$=\nu\int_0^\infty \{f(t)\}^2\,dt. \quad \ldots\ldots\ldots \quad (1)$$

The anode current develops a potential difference across a resistance and the pulse $\delta(t)$ later is transformed into a pulse $f'(t)$ by the pulse shaping network. We therefore derive, following Schottky (1918),

$$\sigma_{\mathrm{I}}^2=\mathrm{R}^2e^2\nu\int_0^\infty \{f'(t)\}^2\,dt$$

$$=\mathrm{R}^2\mathrm{I}_a e\int_0^\infty \{f'(t)\}^2\,dt. \quad \ldots\ldots\ldots \quad (2)$$

Where σ_{I} is the r.m.s. anode current fluctuation. If we refer this to the ionization chamber input and put in the usual modification for space change limitation (North 1940), we obtain

$$\sigma_{\mathrm{Q}}^2=\frac{0{\cdot}12ec^2}{g_m}\,\frac{\mathrm{I}_a}{\mathrm{I}_a+\mathrm{I}_{sg}}\left\{1+\frac{8\mathrm{I}_{sg}}{g_m}\right\}\int_0^\infty \{f'(t)\}^2\,dt\,*$$

(where c is the total input capacity of valve and counter) . (3)

for conventional valves with oxide-coated cathodes and a high amplification factor.

The grid current produces a potential difference across a condenser and the fluctuation becomes

$$\sigma_{\mathrm{Q}}^2=\mathrm{I}_g e\int_0^\infty \{f(t)\}^2\,dt. \quad \ldots\ldots\ldots \quad (4)$$

The background secondary electrons in the counter will produce a background

$$\sigma_{\mathrm{Q}}^2=\mathrm{I}_\gamma(ke)\int_0^\infty \{f(t)\}^2\,dt. \quad \ldots\ldots\ldots \quad (5)$$

Where $\mathrm{I}\gamma$ is the total ionization current due to γ rays and ke is the average charge collected per electron. The noise in the input grid leak is (following Johnson 1925)

$$\sigma_{\mathrm{Q}}^2=2k\mathrm{T}/\mathrm{R}\int_0^\infty \{f(t)\}^2\,dt. \quad \ldots\ldots\ldots \quad (6)$$

which is small compared with the grid current noise if R is large, or more exactly if

$$2k\mathrm{T}/\mathrm{R}\ll(\mathrm{I}_g+k\mathrm{I}_\gamma)e.$$

It can be shown that the anode resistor noise belongs to the same "class" as the anode current noise and is always small in comparison.

---

* All r.m.s. fluctuations are now expressed in terms of a fluctuation of charge on the first valve grid. This is denoted by use of the suffix σ_{Q}.

Flicker noise cannot be classified in either class I. or class II., but is intermediate. The original process is frequency dependent and σ^2 falls inversely as the first power of the frequency. It will only be appreciable for pulse widths greater than 1 millisecond.

Induced grid noise is of the same class (class I.) as anode noise, but is always negligible in comparison. All the usual forms of noise can therefore be represented in the form

$$\sigma_Q^2 = A\int_0^\infty \{f(t)\}^2\,dt + B\int_0^\infty \{f'(t)\}^2\,dt \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where

$$A = (I_g + kI_\gamma)e + 2kT/R, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7a)$$

$$B = \frac{0{\cdot}12eC^2}{g_m}\,\frac{I_a}{I_a + I_{sg}}\{1 + 8I_{sg}/g_m\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7b)$$

For a triode, B reduces to $\dfrac{0{\cdot}12eC^2}{g_m}$.

OPTIMUM PULSE SHAPE.

It is often of interest to obtain the maximum possible signal to noise ratio with no restrictions on the shape of the output pulse. This may be simply derived from the expression above by minimizing σ_Q^2 for a constant peak value of $f(t)$. $f(t)$ must clearly be zero at $t=0$ and $t=\infty$ the first from initial conditions, and the second to make

$$\int_0^\infty \{f(t)\}^2\,dt$$

finite.

Moreover it should be a symmetrical pulse; the integral $\int_0^\infty \{f(t)\}^2\,dt$ can be separated into

$$\int_0^{T_D} \{f(t)\}^2\,dt + \int_{T_D}^\infty \{f(t)\}^2\,dt,$$

where $f(t)$ is a maximum at $t=T_D$. If $f(t)$ is symmetrical the two integrals will be equal; if $f(t)$ is not symmetrical and

$$\int_0^{T_D} \{f(t)\}^2\,dt > \int_{T_D}^\infty \{f(t)\}^2\,dt,$$

then σ will be made smaller by altering $f(t)$ between 0 and T to make the pulse symmetrical.

If $f(t)$ is a function of (t/τ) alone where τ is some variable, then the best choice of τ is given when the two integrals which add to produce σ^2 are equal, *i. e.* when

$$A\int_0^\infty \{f(t)\}^2\,dt = B\int_0^\infty \{f'(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If we can choose the form of the function $f(t)$ then the best function is given by

$$f(t)=\exp\{-|t-T_D|/\tau\} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (9)$$

and $\tau^2=B/A$.
This follows by minimizing

$$\int_{T_D}^{\infty} A\{f(t)\}^2\,dt+\int_{T_D}^{\infty} B\{f'(t)\}^2\,dt=\sigma^2/2$$

with the auxiliary conditions, that at

$$t=\infty \quad f(t)=0,$$
$$t=T_D \quad f(t)=1,$$

this is normalizing the noise to unit pulse height. Then using the Calculus of Variations we obtain

$$2f(t)=\frac{B}{A}\frac{d}{dt}\{2f'(t)\} \quad \text{(Euler's formula)},$$

$$f(t)=\frac{B}{A}f''(t),$$

whence $f(t)=\exp\{-(t-T_D)/\tau\}$ where for $t>T_D$ and $f(t)=\exp\{(t-T_D)/\tau\}$ for $t<T_D$ by the symmetry property discussed above. There is a discontinuity in $f'(t)$ at $t=T_D$ but no discontintinuity in $\{f'(t)\}^2$.

Clearly

$$A\int_0^{\infty}\{f(t)\}^2\,dt=B\int_0^{\infty}\{f'(t)\}^2\,dt$$

in agreement with formula (8).

From a physical point of view, it is of interest to see why this pulse shape is the best. The sharp pointed peak will not give use to excessive noise, since the extra height adds little to the area under the curve of $\{f(t)\}^2$, whereas it adds considerably to the signal. This pointed peak is the great advantage.

COMPARISON OF PULSE SHAPES.

It remains to be seen how much improvement can be obtained in practice by using this pulse shape in preference to other pulse shapes; in order to do this it is necessary to consider the response of commonly used amplifier networks to a pulse $H(t)$ or $\delta(t)$ to obtain the shape $f(t)$ and $f'(t)$.

The response of the amplifier to a function $H(t)$ or $\delta(t)$ may be readily found by the method of the Laplace Transformation.

$$F(p)=\int_0^{\infty} e^{-pt}f(t)\,dt. \quad . \; . \; . \; . \; . \; . \; . \; . \quad (10)$$

The transform of $\delta(t)$ is 1 and of $H(t)$ is $1/p$.

The system function for the amplifier $G(p)$ is written down by replacing the differential operator d/dt by p. Then the response to a pulse $\delta(t)$ is the inverse transform of $G(p)$ and the response to a pulse $H(t)$ is the inverse transform of $G(p)/p$. The method is discussed by Elmore (1948) and Carslaw and Jaeger (1941). The method is here applied to obtain the signal to noise ratio for a variety of pulse-shaping circuits.

(1) Single differentiating circuit and single integrating circuit.

The system function $G(p)=\tau_2/(1+\tau_1 p)(1+\tau_2 p)$

$$f(t)=(\tau_2/\tau_1)^{1/(1-\tau_2/\tau_1)}\{\exp(t/\tau_2)-\exp(t/\tau_1)\}(\tau_2/\tau_2-\tau_1) \quad . \quad . \quad (11)$$

when normalized to unity maximum amplitude (Elmore 1948).

σ^2 is then a minimum for $\tau_1=\tau_2=\tau=\surd(B/A)$ and

$$\sigma^2=\surd(AB),\ \frac{e^2}{2}=3{\cdot}7\ \surd(AB). \quad . \quad . \quad . \quad . \quad . \quad (12)$$

A similar response is obtained from using a critically damped tuned circuit with $\tau^2=LC$.

(2) A shorted delay line in conjunction with a single integrating circuit gives optimum signal to noise ratio if the delay line time (for twice the line length) is equal to approximately $1{\cdot}2\tau$ where τ is the integrating time constant, and

$$\tau=\surd(B/A).$$

Then

$$\sigma^2=2{\cdot}5\ \surd(AB). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

(3) A large number of integrating circuits limit the response of a multi-stage amplifier. After a large number of such circuits the response to a pulse $\delta(t)$ tends to the shape

$$f(t)=\exp\{-(t-T_D)^2/\tau^2\} \quad \text{(Elmore 1948),} \quad . \quad . \quad . \quad (14)$$

where $T_D=n\tau_1$ and $\tau=\tau_1\surd(n)$.

Where n is the number of stages. When n is large, this is also the response to a function $H(t)$ if this has been differentiated with a time constant small compared with τ. Pulse shaping with a shorted delay line and a large number of integrating circuits will also give approximately the same shape. Then

$$\sigma^2=2{\cdot}5\surd AB \text{ with } \tau=\surd A/B. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The optimum case with

$$f(t)=\exp\{-\,|\,t-T_D\,|\,/\tau\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15a)$$

with $\tau=\surd A/B$ gives

$$\sigma^2=2\surd AB. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

These figures differ from those derived by Elmore (1948). Elmore defines a pulse width

$$W^2=\frac{2\pi\int_0^\infty (t-t_0)^2 f(t)\,dt}{\int_0^\infty f(t)\,dt}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (17)$$

where

$$t_0=\frac{\int_0^\infty tf(t)\,dt}{\int_0^\infty f(t)\,dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (17a)$$

and derives the minimum noise—neglecting the grid noise (*i. e.* putting $A=0$) for a constant value of W. It is of interest to compute W for our various pulse shaping networks when the parameter t_0 has been chosen to make σ^2 a minimum.

For the optimum case $W^2= 4\pi\frac{A}{B}$.

For the Gaussian case $W^2= \pi\frac{A}{B}$.

For the RC–RC case $W^2=12\pi\frac{A}{B}$.

For the RC–Delay line case $W^2= 4\pi\frac{A}{B}$.

Thus the Gaussian case has the advantage of giving a smaller pulse width—according to Elmore's definition. This is in fact because there is less of a " tail " to the pulse.

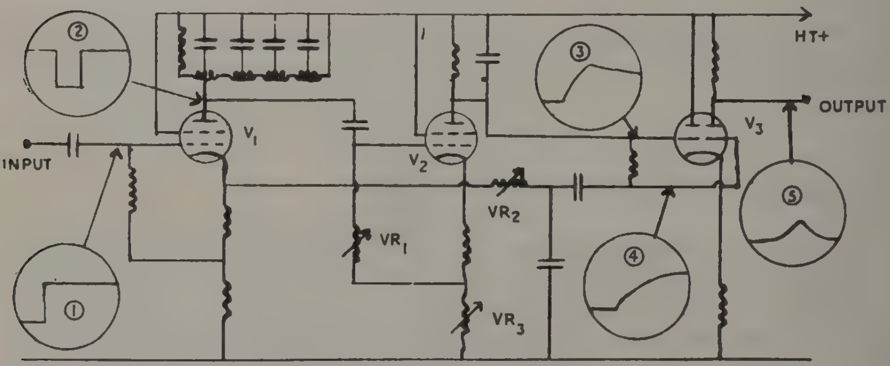
For most purposes it will not be necessary to produce a circuit to shape the pulse to the optimum. The improvement over a Gaussian or RC–Delay line case is only 15 per cent. For some experiments, however, this extra 15 per cent may make all the difference. This arises when low energy protons are to be counted in the presence of intense γ radiation. (Wilson, Collie and Halban 1949). A circuit has been devised (fig. (1)) which will shape the pulses to something which closely approaches the optimum shape. It is necessary, for such a circuit, that only *linear* networks be used ; otherwise, although a pulse $H(t)$ may be shaped to a pulse, $f(t)$ a pulse $2H(t)$ will not be shaped to a pulse $2f(t)$.

Briefly the operation of the circuit is as follows. The input pulse (1) is split into two parts. One half is converted to a square pulse (2) and then into a nearly linear voltage rise (3). The other half is converted into an exponential rise (4). The first shape is controlled by the delay length and VR1, the second shape by VR2. The relative amplitudes are controlled by the feedback resistor VR3, so that the initial slopes of both waveforms (3) and (4) are the same. The double triode V_3 is connected so that the difference between the two waveforms is taken as the output. The pulse shape after the peak is exactly the optimum shape, but that before the peak rises too quickly (theoretically it should begin to rise an infinite sime before the peak). With suitable choice of constants

$$\sigma^2=2{\cdot}1\sqrt{AB}. \qquad (18)$$

This computation made above of the optimum pulse shaping network assumes that the valve anode and grid noise are independent and constant. If the valve operating conditions are altered, the grid current noises can be reduced, at the expense of the anode current noise, by reducing the HT voltage (and hence the space current) and also by slightly reducing the filament voltage. The minimum noise is given when AB is a minimum (provided that the amplifier time constants are readjusted to the optimum values). If only valve noise and γ background are important this means $(kI_\gamma+I_g)/g_m$ must be a minimum. This is a slight extension of a formula of Gillespie (1948). This assumes that the anode current noise is completely space charge limited, and therefore only holds for a triode.

Fig. 1.



Pulse shaping circuit.

If two different valves are to be compared, then the input capacity alters. A "figure of merit" for a valve in any given application is therefore

$$g_m/C^2(I_g+kI_\gamma). \qquad (19)$$

Where C is the total input capacity of valve and counter (but *not* the feedback capacity due to Miller effect).

Counting of Noise Peaks.

Before it is possible to compare these results with practice it is necessary to consider the rate of counting of noise peaks.

The number of times that the noise waveform passes through a voltage V with positive slope is

$$[\exp\{-V^2/2\overline{V_0^2}\}]\{\tfrac{1}{2}\text{ expected number of zeros of the noise waveform}\}$$

(Rice 1944, 1945, Middleton 1948). . . . . . . (20)

This is clearly the number of times the discriminator triggers.

The expected number of zeros per second is

$$2\left[\frac{\int_0^\infty f^2 W(f)\,df}{\int_0^\infty W(f)\,df}\right]. \qquad \ldots \ldots \quad (21)$$

Where $W(f)\,df$ is the noise power at frequencies between f and $f+df$.

For the cases considered this approximates to $0{\cdot}6/\tau$ where $\tau=\sqrt{B/A}$. (The case of an integrating R–C circuit as the upper frequency cut off, gives a divergent integral; this is discussed by Rice but will not arise in physically realizable circuits.)

For most of our cases $\tau=10^{-6}$ secs.

Then the number of zeros is $0{\cdot}6\times10^{6}$/second.

Then there will be less than 1 count per minute on noise peaks if $V/V_0>6$, *i. e.* if the signal is greater than 6 times the r.m.s. noise.

PRACTICAL AMPLIFIER CIRCUIT.

The above analysis has been carried out in order to design a counter-amplifier system to count low energy protons in the presence of intense γ radiation for work on photodisintegration of deuterium (Wilson, Collie and Halban, 1949). The valve chosen for first valve is an American 6AK5 connected as a triode by strapping screen and anode. This reduces the anode current shot noise. The total input capacity of valve and counter is 13 pF. The grid leak is made up only of the valve and insulator leakages and is so high that no noise arises.

The 6AK5 valves used were well aged, and then selected; the grid current was then smaller than any other type tested; the Mullard EF37, although it has a low grid current, gives more anode current noise, and the EC91, or EF91, although they give little anode current noise have a higher input capacity and more grid current noise.

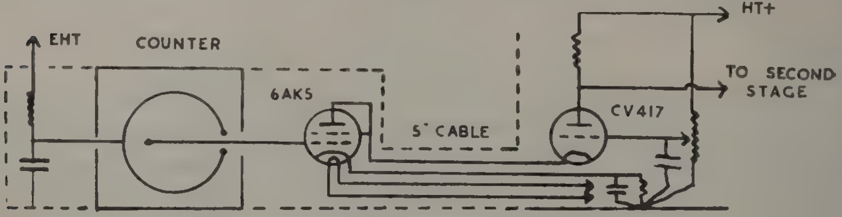
Johnson and Johnson (1936) and Parsegian (1946) recommend pentodes for first valve to avoid the "Miller effect" feedback. This feedback *does not* affect the signal to noise ratio, but it affects the gain and may result in accentuation of second valve noise. This can be obviated by a compensating positive feedback from the second valve anode to the first valve grid (Schultz 1946). We have, however, inserted an extra valve in the "cascode" connection (Hunt and Hickman 1939, Wallman, MacNee and Gadsden 1948). The anode of the first valve feeds directly to the cathode of the second which represents a sufficiently low impedance (100 Ω) to prevent appreciable amplification and feedback by the first valve. Yet the second valve noise is reduced by the high impedance of the first valve in the cathode. This circuit also enables the first valve to be placed away from the subsequent valves—separated by a yard or so of cable (fig. 2).

After amplification by a feedback ring a 100 Ω line is taken to the main amplifier where the pulse shaping networks are situated. A shorted delay line and resistance capacitance integration are normally used. It is possible, with selected valves, to make grid and anode current noise equal with $\tau=0{\cdot}5\times10^{-6}$ secs. with $g_m=4$ mA/V. Under these conditions, the equivalent number of ion pairs for the noise$=\sigma/c=300$.

Is is therefore possible to count pulses of ionization of 2000 ions pairs above background, or 70 keV.

When a γ ray source is brought near the high pressure counter used, the value of τ is reduced, to $0{\cdot}3\times10^{-6}$ secs. (below which loss of pulse height might occur) and it is only possible to count down to about 90 keV. These values of the background, and the rate of counting of noise pulses (equation (20)) are fully borne out experimentally, confirming that there are no other forms of background than the statistical ones considered here.

Fig. 2.



Amplifier input circuit.

ACKNOWLEDGMENTS.

The thanks of the author are due to Mr. Gillespie and Dr. D. K. C. MacDonald for their advice on noise problems ; to Dr. T. W. Chaundy for some mathematical assistance, to Mr. C. H. Collie for suggesting the problem and for his continuous encouragement, and to Mr. G. R. Bishop for his help in the assembly of the apparatus.

REFERENCES.

CAMPBELL, N., 1909, *Proc. Camb. Phil. Soc.*, **15**, 117.
CARSLAW, H. S., and JAEGER, J. C., 1941, *Operational Methods in Applied Mathematics*. O.U.P.
ELMORE, W. C., 1948, *J. App. Phys.*, **18**, 55 ; *Ibid.*, *Nucleonics*, **2**, 23.
GILLESPIE, A. B., 1948, *A.E.R.E. Report*.
HUNT, F. V. and HICKMAN, R. W., 1939, *Rev. Sci. Inst.*, **10**, 6.
JOHNSON, J. B., 1925, *Phys. Rev.*, **26**, 71.
JOHNSON, E. A., and JOHNSON, A. G., 1936, *Phys. Rev.*, **50**, 175.
MIDDLETON, D. O., 1948, *J. App. Phys.*, **19**, 817.
NORTH, D. O., 1940, *R.C.A. Rev.*, **5**, 244.
PARSEGIAN, V. L., 1946, *Rev. Sci. Inst.*, **17**, 39.
RICE, S. O., 1944, *Bell System Tech. Jnl.*, **23**, 282 ; 1945, *Ibid.*, **24**, 146.
SCHOTTKY, W. 1918, *Ann. Phys. Lpz.*, **57**, 541.
SCHULTZ, H. L., 1946, *Phys. Rev.*, **69**, 689.
WALLMAN, H., MACNEE, A. B., and GADSDEN, C. B., 1948, *Proc. Inst. Rad. Eng.*, **36**, 700.
WILSON, R., COLLIE, C. H., and HALBAN, H. H., 1949, *Nature*, **163**, 245.

VI. *A Radio Echo Method for the Measurement of the Heights of the Reflecting Points of Meteor Trails.*

By J. A. Clegg and I. A. Davidson,
Physical Laboratories, The University, Manchester *

[Received September 27, 1949.]

Abstract.

This paper describes a method of finding the heights of occurrence of meteor trails from the radio echoes obtained by broadside reflection on a frequency of 60 Mc/s. The range of the reflecting point from the observing station is measured by the radio pulse technique and its elevation is determined by comparison of the signals received in two aerials mounted at different heights above a horizontal reflecting surface. Results obtained by this method during the Quadrantid shower of 1949 are in good agreement with those of visual observers.

§1. Introduction.

It is well known that a meteor, in its passage through the atmosphere. produces a column of ionization which can be detected by radio echo methods. Such methods have been developed independently by Hey and Stewart (1947) and Lovell (1948) in Great Britain ; by McKinley and Millman (1949) Canada ; and by Manning, Villard and Peterson (1949) United States.

It has been shown by Hey and Stewart (1947), and by Lovell, Banwell and Clegg (1947), that radio waves of sufficiently high frequency (60 to 75 Mc/s.) undergo specular reflection from the ionized column, the echo being returned from a small element of the trail situated close to the foot of the perpendicular from the station. A study of the durations of such transient echoes presents a number of interesting physical problems, and precise determinations of the heights of the reflecting points is necessary before progress can be made in their solution.

Although the radio echo technique has been used previously to find the heights of occurrence of meteors in the atmosphere, the early methods depended on combined radio and visual observations (Prentice, Lovell and Banwell, 1947, Manning, Villard and Peterson, 1949), and were, moreover, not capable of high accuracy. A more recent method of improved accuracy described by Millman and McKinley (1949) can only be used on low frequencies (30 Mc/s.) where the reflection becomes non-specular.

This paper describes a height finding system which operates on a considerably higher frequency, where the broadside reflection condition invariably holds. It is based on the spaced aerial method (Taylor and Westcott, 1946), whereby the range and elevation of the reflecting point from a single observing station are determined by radio measurements alone. The method was used to find the heights of 28 meteors during the Quadrantid shower of 1949, and has since been employed during the daytime showers of 1949 May to July.

---

* Communicated by Dr. A. C. B. Lovell.

§2. PRINCIPLE OF THE METHOD.

The apparatus operates on a frequency of 60 Mc/s. The range of the reflecting point of the meteor trail is measured by the radio pulse method, and its elevation is found by comparing the signals received in two aerials mounted at different heights above a plane, horizontal, perfectly reflecting surface.

For a scattering source of small dimensions at elevation ψ from the station, the ratio, r, of the signals received in two such aerials is equal to the ratio of their effective power gains, $G_1(\psi)$ and $G_2(\psi)$, in this direction, and as these functions are calculable from the known parameters of the system, the relationship between ψ and r can be readily established. For a meteor trail, which subtends a finite angle at the station, the echoes are due to the integrated effect of the radiation scattered from all parts of the ionized column, and the evaluation of ψ in terms of r becomes more difficult. It can, however, be shown as follows, that provided the aerials are not too highly beamed, the amplitude of the echo received in either is proportional to its power gain in the direction of the reflecting point.

Consider first the case of scattering from a trail whose reflecting point lies close to the axis of a wide beam. The overall power gain G, of the combined transmitting and receiving aerial systems, can then be assumed to be constant over the first few Fresnel zones. It has been shown by Lovell and Clegg (1948) that under these circumstances the amplitude of the voltage fluctuations set up in the first stage of the receiver by scattering from a section of the trail lying symmetrically about the reflecting point and subtending an angle 2θ at the station is

$$V=k(\lambda/R)^{3/2}\alpha G\left\{\int_{-v}^{+v}\cos \pi/2v^2\,dv+i\int_{-v}^{+v}\sin \pi/2v^2\,dv\right\}. \quad . \quad . \quad (1)$$

Where k is a constant depending on the parameters of the apparatus, λ is the wavelength, R is the range of the reflecting point, α is the electron line density, and $v=2\theta\sqrt{(R/\lambda)}$. The amplitude of the signal received from the whole trail is obtained by taking the limiting values of the Fresnel integrals as v increases indefinitely.

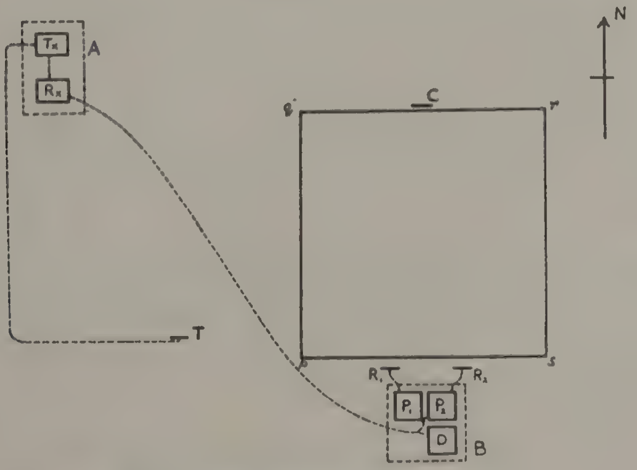
For accurate elevation finding it is necessary to make observations in directions where the gain of the spaced aerials varies rapidly with θ, so that to derive an expression for the amplitude of the signal received in either it is necessary to include G under the integral sign in (1). The value of V will therefore depend in general on the radiation pattern of the aerial. However, over a small angular range on either side of the reflecting point, G will approximate closely to a linear function of θ, and may be written

$$G_0 + \theta \partial G/\partial \theta = G_0 + v \partial G/\partial v.$$

Where G_0 is the value of the gain in the direction of the reflecting point. If the width of the beam is sufficiently great in comparison with the angle subtended by the first few Fresnel zones this value may be substituted in (1), which then becomes

$$V = k(\lambda/R)^{3/2}\alpha\left\{G_0\left[\int_{-v}^{+v}\cos(\pi v^2/2)\,dv + i\int_{-v}^{+v}\sin(\pi v^2/2)\,dv\right]\right.$$
$$\left. + \partial G/\partial v\left[\int_{-v}^{+v} v\cos(\pi v^2/2)\,dv + i\int_{-v}^{+v} v\sin(\pi v^2/2)\,dv\right]\right\}.$$

Fig. 1.



Plan of the apparatus (not to scale).

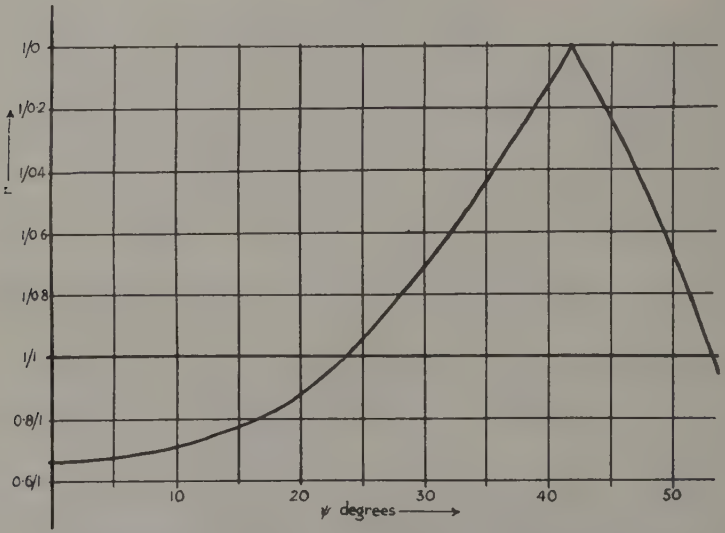
The expressions under the integral signs in the second term of the right-hand side of this equation are odd functions of v, and the integrals vanish identically. The equation therefore reduces to (1), the amplitude of the echo is proportional to G_0, and the ratio of the signals received in the two aerials of a spaced system is equal to the ratio of their gains in this direction.

§3. APPARATUS.

A plan of the apparatus is shown in fig. 1. The hut A contains a transmitter and receiver of conventional design. The transmitter works on a frequency of 60 Mc/s. ($\lambda = 5$ metres), and radiates pulses of 8 μ/sec duration and peak power 100 kw. The amplification of the receiver is linear, and its sensitivity is such that a signal of 2 μV. at the input doubles the signal to noise ratio. The transmitting array, T, floodlights the northern quadrisphere of the sky. It consists of a horizontal folded half-wave aerial at a height of $\frac{5}{8}\lambda$ above the ground, with a reflector adjusted to give minimum backward radiation.

The two receiving aerials R_1 and R_2 are similar to the transmitting array, and are mounted on posts 20 ft. apart at a distance of 200 ft. from the hut A. In front of these aerials is a plane macadam square, *pqrs*, with sides 100 ft. in length which can be covered with wire mesh to form a perfectly reflecting surface. The heights of the receiving arrays above the square can be varied from 3 ft. to 18 ft., but they are normally maintained at $\frac{1}{2}\lambda$ and $\frac{3}{4}\lambda$. The aerials are connected to the two stage preamplifiers P_1 and P_2 housed in the small hut B, and the signals pass from here through a common coaxial cable to the receiver. The preamplifiers are suppressed alternately at a frequency of 25 c/s. by multivibrator pulses supplied from the receiver. The same multivibrator pulses are applied to the x-plates of the A scope display tube, so that the echoes from the two aerials are displayed side by side. Their relative amplitudes are compared visually.

Fig. 2.



Theoretical relation between ratio r and elevation ψ.

For purposes of calibration a fourth aerial C is mounted at a height of 20 ft. in the centre of the side of the square opposite to the receiving aerials. A duplicate display tube D with its associated time base and power circuits is provided in the hut B. Before each set of readings the aerials R_1 and R_2 are brought to the same height above the reflecting surface and a signal generator is fed into the aerial C. The gains of the preamplifiers are then adjusted until the signals received in R_1 and R_2

are of the same amplitude, and the aerials are then returned to their original heights of $\frac{1}{2}\lambda$ and $\frac{3}{4}\lambda$.

Fig. 2 shows the ratio r of the effective aerial gains plotted as a function of the elevation, ψ. As the aerials fulfil the conditions required by §2, this curve can be used to obtain the elevation of the reflecting point of a meteor trail from the observed value of r.

§4. LIMITATIONS AND ERRORS.

(i) *Error in Measurement of the Ratio r.*

With the apparatus in its present form the most serious inaccuracy arises from the uncertainty in the estimation of the ratio r. To find the magnitude of the error involved artificial echoes were displayed on the cathode ray tube, and the apparent visual ratio was compared with that obtained by photographic measurement. The results showed that when the amplitude of the larger echo exceeds that of the noise by a factor of 2·5 or more, the ratio can always be measured to within 1 part in 10, and that for signals differing in amplitude by less than 20 per cent accuracy of 1 part in 20 is possible. For values of ψ greater than 20° the consequent error in the height is never greater than ± 4 per cent.

(ii) *Ambiguity.*

The vertical polar diagram of the higher aerial has two lobes with a zero at 41·8°, so that there are two values of ψ corresponding to each value of r. One of these usually gives an improbable value for the height H, and can be rejected. For a trail at elevations between 33° and 48°, however, there are invariably two possible values of ψ, and the method cannot be used.

(iii) *Error in Range Measurement.*

With the present apparatus ranges can be measured to within ± 2 km. This corresponds to an error in height measurement of ± 1 km.

(iv) *Imperfections in the Reflecting Surface.*

The method depends on the fact that the reflecting surface acts as a plane mirror with a reflecting coefficient of unity. It is well known that for this condition to obtain the surface must behave as a perfect plane reflector over an area comparable with that of the first Fresnel zone. At the lower limit of elevation (20°) the distance of the geometrical reflecting point from the higher aerial is less than 50 ft., but the first Fresnel zone extends outwards to a distance of 180 ft., and a considerable part of its area lies outside the surface *pqrs* in fig. 1. It is therefore important that the ground shall have good reflecting properties for some distance beyond the edges of the square.

For 120 ft. beyond the sides *pq*, *qr*, and *rs*, the earth is at the same level as the prepared surface, and is plane to within $\lambda/16$. The average values given by Burrows (1937) for the dielectric constant and the electrical conductivity of soil are 10 and 5×10^7 e.s.u. respectively. This implies that for waves of frequencies of 60 Mc/s. its refractive index, n, is equal to $10-i5/3$. Substituting this value in Fresnel's equation we find that for horizontally polarized waves the reflexion coefficient at a glancing angle ψ is given by

$$\rho=\frac{\sin\psi-\sqrt{(10-i\,5/3\cos^2\psi)}}{\sin\psi+\sqrt{(10+i\,5/3\cos^2\psi)}}.$$

In directions within 30° of true north the geometrical reflecting point lies outside the area *pqrs* only when ψ is less than 7°, or when ρ is greater than 0·94. The modification in the vertical polar diagram due to this small departure from perfectly reflecting conditions is inappreciable at elevations greater than 20°. We may therefore assume that over the range of elevations with which we are concerned the curve in fig. 2 is unaffected by the limited area of the prepared surface.

(v) *Coupling between the two receiving aerial systems.*

To measure the degree of coupling between the two receiving aerials, a signal generator was fed into the aerial C (fig. 1), and the signal from the aerial R_2 was permanently suppressed. The amptitude of the signal received in R_1 was measured by a microammeter connected across the output of the receiver. The aerial R_2, which was initially at a height of 3 ft., was raised slowly to a height of 18 ft., and was then disconnected. During this operation there was no perceptible change in the reading of the meter. The coupling between the two aerials must therefore be negligible.

(vi) *Change of aerial impedance with height above the ground.*

The method of calibration described in §3 assumes that the impedance of the receiving aerials is independent of their height above the reflecting surface. To measure the variation in impedance with height a signal was radiated from the aerial C and received in the aerial R_1 in the way described in (v) above. It can readily be shown that if the impedance of R_1 is independent of its height, h, the strength of the signal received will be proportional to $\sin\kappa h$, where κ is a constant. If, however, the mutual impedance between the aerial and its image is appreciable, the signal will show fluctuations in amplitude about the sine curve as h is varied.

On varying h from 3 ft. to 18 ft. these fluctuations were observed, but their amplitude corresponded to less than 0·5 per cent of the total amplitude of the signal. The change in impedance with height is therefore trivial.

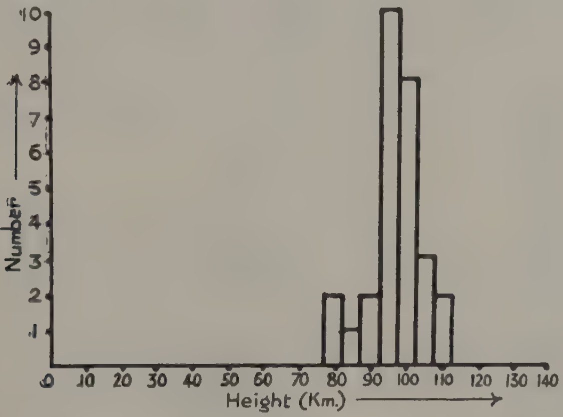
§5. RESULTS OBTAINED DURING THE QUADRANTID SHOWER OF 1949.

This method of height measurement was tested during the Quadrantid meteor shower of 1949. The periods of observation were from January 3, $03^h\,08^m$ to $04^h\,40^m$, January 3, $13^h\,00^m$ to $13^h\,49^m$, and January 4,

$03^h\ 08^m$ to $04^h\ 30^m$. At these times the ground was covered with a layer of wet snow to a depth of 5 in. To find the effect of this snow on the reflecting properties of the surface, the experiment described in §4, section (vi) was repeated during the early morning of January 3. No deviation from perfectly reflecting conditions was detected.

The mean echo rate during the periods of observation was 14 per hour. Previous experience with similar apparatus of comparable sensitivity has shown that the rate when no shower is active is two per hour. It may therefore be assumed that the majority of the echoes observed were due to meteors from the Quadrantid radiant. Out of a total of 41 echoes two were at elevations less than 25°, nine gave ambiguous heights, and in two cases there was uncertainty in the measurement of the ratio. The remaining 28 measurements gave heights between 80 km. and 107 km., with a probable error of less than 5 km. in each case. The height distribution is shown in fig. 3.

Fig. 3.



Height distribution for Quadrantid shower, 1949.

§6. RESULTS OBTAINED DURING THE DAYTIME SHOWERS OF 1949, MAY TO JULY.

The method has recently been used to measure the heights of meteors of the summer daytime streams of 1949. The apparatus was slightly modified to give greater accuracy of range measurement, the pulse recurrence frequency being increased to 150 c/s., and the time base expanded to cover only 250 km. During the period 1949 May to July watches were maintained for a total period of 8·26 hours, and 239 echoes were observed. Of these 166 gave unambiguous heights. The height distribution is shown in fig. 4.

§7. DISCUSSION OF RESULTS.

The results obtained by this method show that it is capable of giving values for the heights of approximately 70 per cent of the total number of meteors observed.

The beginning and end heights obtained visually by Prentice for trails of the Quadrantid meteors are shown in Table I.

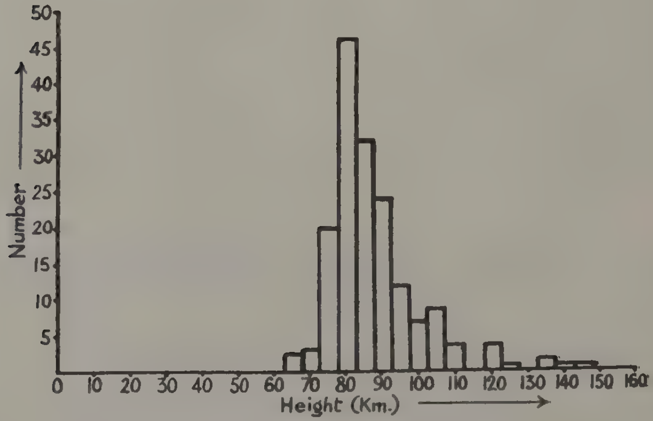
TABLE I.

Heights of Occurrence of Quadrantids obtained from visual measurement.

Visual Magnitude	No. of Meteors Observed	Mean Beginning Height (H_1)	Mean End Height (H_2)
0	1	108 km.	84 km.
2	5	105 km.	87 km.
4	28	103 km.	91 km.
6	140	101 km.	94 km.

Mean Height of Occurrence $\frac{(H_1+H_2)}{2}=97{\cdot}8$ km.

Fig. 4.



Height distribution for daytime showers, 1949.

It is evident that the heights shown in this table are confined within somewhat narrower limits than those obtained by the radio echo method, but this is to be expected, since the values shown here represent the means of a relatively large number of observations. The value of the mean height given by Prentice is in good agreement with that obtained by the radio technique.

Although it is not possible to obtain a direct comparison between the results obtained by this and other methods for meteors of the daytime streams, the height distribution shown in fig. 4 is of a form which would be expected for meteors of medium velocity.

§8. CONCLUSION.

We may conclude, from the results listed above, that the method in its present form is capable of measuring the heights of the reflecting points of meteor trails to within ±4 km. The discussion in §4 indicates that by far the most serious source of error is the measurement of the ratio of the echoes. This may be eliminated by the addition of automatic photographic recording, which will also reduce the range error. It is estimated that with the addition of this facility, a probable error of ±1 km. will be attainable.

§9. ACKNOWLEDGMENTS.

The authors wish to thank Dr. A. C. B. Lovell, the Director of the Jodrell Bank Experimental Station, for his help and encouragement. The cost of the apparatus, including the construction of the flat reflecting surface, has been met from the grant supplied by the Department of Scientific and Industrial Research for the development of the experimental work at Jodrell Bank. One of the authors (I. A. D.) also wishes to thank the Department of Scientific and Industrial Research for the award of a Maintenance Grant which has enabled him to carry out this work.

REFERENCES.

BURROWS, C. R., 1937, *Bell System Tech. J.*, **16**, 45.
HEY, J. S., and STEWART, G. S., 1947, *Proc. Phys. Soc.*, **59**, 858.
LOVELL, A. C. B., 1948, *Phys. Soc. Report on Progress in Physics*, **11**, 415.
LOVELL, A. C. B., BANWELL, C. J., and CLEGG, J. A., 1947, *M. N. Roy. Astr. Soc.*, **107**, 164.
LOVELL, A. C. B., and CLEGG, J. A., 1948, *Proc. Phys. Soc.*, **60**, 491.
MCKINLEY, D. W. R., and MILLMAN, P. M., 1949, *Proc. I.R.E.*, **37**, 364.
MANNING, L. A., VILLARD, O. G., and PETERSON, A. M., 1949, *J. Appl. Phys.*, **20**, 475.
MILLMAN, P. M., and MCKINLEY, D. W. R., 1949, *Sky and Telescope*, **8**, No. 5.
PRENTICE, J. P. M., Privately communicated.
PRENTICE, J. P. M., LOVELL, A. C. B., and BANWELL, C. J., 1947, *M. N. Roy. Astr. Soc.*, **107**, 155.
TAYLOR, D., and WESTCOTT, C. H., *J. Inst. Elect. Engrs.*, 1946, **93**, part IIIA, 588.

VII. *On the Mass of Charged Particles of the Cosmic Radiation.*

By C. Franzinetti, D. in F.,
H. H. Wills Physical Laboratory, University of Bristol*.

[Received October 31, 1949.]

[Plates I. & II.]

Summary.

The mass of charged particles, produced by cosmic radiation at an altitude of 11,000 ft., and brought to rest in photographic emulsions, has been determinedby 'magnetic-deflection' experiments. The method is applicable to particles, of any type, of which the lifetime is greater than 10^{-10} sec.

The results show that at an altitude of 11,000 feet the particles are composed of different types in the following proportions :—

μ^+-particles, $8{\cdot}5 \pm 1{\cdot}5$ per cent ; μ^--particles, $8 \pm 2{\cdot}4$ per cent ; π^+-particles, $4{\cdot}0 \pm 1{\cdot}0$ per cent ; π^--particles, $5{\cdot}4 \pm 2{\cdot}0$ per cent ; Protons, 48 ± 4 per cent ; deuterons, 16 ± 2 per cent ; more massive positive particles, 11 ± 2 per cent.

The observations on the tracks of 360 particles give no evidence for the existence of other kinds of particles sufficiently stable to be detected by this method ; and they show that if such particles exist, they are very rare.

§ 1. Introduction.

Some time ago it was pointed out by Powell and Rosenblum (1948) that with a suitable arrangement of photographic plates placed in a magnetic field it is possible to measure both the momentum and the residual range of a charged particle, and thence to determine its mass. It was proposed that two plates should be placed parallel to one another and perpendicular to a magnetic field, the two emulsions being face to face. A charged particle leaving one of the emulsions at a small glancing angle, ϕ, and with sufficient energy to reach the opposite emulsion, will be recorded in both plates.

Owing to the magnetic field, the particle will move, in the space between the plates, along a spiral path lying on a circular cylinder of which the radius, ρ, will depend on the component of the momentum of the particle

---

[*] Communicated by Professor C. F. Powell.

in the plane perpendicular to the field. From the angle between the directions of motion of the particle, at the points of exit and entry at the two emulsions, ρ can be determined. If the particle stops in the second emulsion, its residual range, r, can also be measured. Inserting the appropriate values of ρ and r in the equations for the momentum, p, and the energy E ; viz. :

$$p=\left(\frac{eZ}{c}\right)\cdot\frac{H\rho}{\sin\phi} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$E=f(r, m, Z) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

the mass of the particle, m, can be determined if its charge, Ze, is known.

It was suggested by Powell and Rosenblum that this method would give more precise values for the masses of the π and μ-mesons than can be obtained either by means of grain-counts or by observations on the scattering of the particles. At the time, such improved measurements were of great interest because of their bearing on the nature of the neutral particle emitted during the decay of the π- meson, but shortly afterwards the announcement was made of the artificial production of π-mesons at Berkeley. It seemed clear that these experiments would give much more favourable conditions for the determination of the mass of the particles, and the cosmic-ray experiments were therefore temporarily abandoned.

Interest in the method has been recently revived by reports, from several laboratories, of observations which have been taken to indicate the existence of other types of mesons in addition to the π- and μ-particles. Further, some experimenters (Alichanian *et al.* 1948) have suggested that these other types occur with a frequency comparable with that of the π- and μ-mesons. The question of the existence or non-existence of such particles is of such decisive importance for the development of nuclear physics. It was therefore decided to proceed with the magnetic deflection experiments, both because they appeared to be capable of greater precision than other methods employing photographic plates, and because they have the advantage that any secondary effects associated with the arrest of the particles in the emulsion can be observed. The second of these features is of particular importance because it is reasonable to assume that any other types of mesons must, like the π- and μ-particles, be unstable ; and any positive particles should therefore give distinctive evidence of spontaneous decay.

A further important advantage of the method is that the time of flight of a particle in the apparatus is only of the order of 10^{-10} sec., a period much shorter than that necessary in experiments employing Wilson chambers and counters. It is therefore possible, in principle, to determine the mass of particles, by the photographic method, even although their existence is so transient that they would decay " in flight " in traversing apparatus employing other methods of detection.

The present paper describes an application of the method to the determination of the mass of charged particles observed in photographic plates exposed at the Jungfraujoch High Altitude Research Station.

§ 2. DETAILS OF EXPERIMENTAL PROCEDURE.

Apparatus. The experimental arrangement employed is shown schematically in fig. 1(*a*). The plates were carried by a suitable plate-holder, (B), which was placed between the pole-pieces of an electro-magnet. A metal housing, (A), was fitted around the pole-pieces in order to maintain the plate-holder in its correct position during the exposure, and to make it possible to evacuate the space between the plates if it was found to be desirable.

The plates employed were coated with Kodak NT4, or Ilford G5 "nuclear research" emulsions, 50μ, 100μ or 200μ thick, and with linear dimensions $3\frac{1}{4} \times 3\frac{1}{4}$ inches. Emulsions thicker than 200μ have not been found convenient for the investigation, because of the existence of distortion in the processed plates. The most important advantage of using thicker emulsions is that there is an increase in the number of tracks which stop in the emulsion, and which are therefore measurable. The results given in a later paragraph show, however, that the trajectory of a proton with a residual range of 1000μ has a radius of curvature of $\sim$20 cm., in the conditions of the experiment. The corresponding value of the deflection, if the path in the air-gap is $\sim$2 cm., is only 6°, and such a small deflection does not allow a precise measurement of the mass of the particle. The gain in the number of tracks obtained in a given exposure with thick emulsion would therefore be balanced by a reduction in the average value of the precision of the determinations of mass. These considerations do not apply if the observations are confined to the particles of small mass, such as π- or μ-mesons, which are more deflected. It was important, however, for the purposes of the present investigation, to make measurements on particles of all types.

Any small movement of the plates during exposure would have made it very difficult to correlate the two tracks produced by a single particle and would have greatly reduced the accuracy of the measurements. The plates were therefore attached to the holder with "Kodaflat" liquid rubber. The rubber was poured on the frame of the holder in sufficient quantity to make a layer 300μ thick, and was allowed to dry for 12 hours. The plates were then pressed against the adhesive surface, and all possibility of their relative movement was excluded.

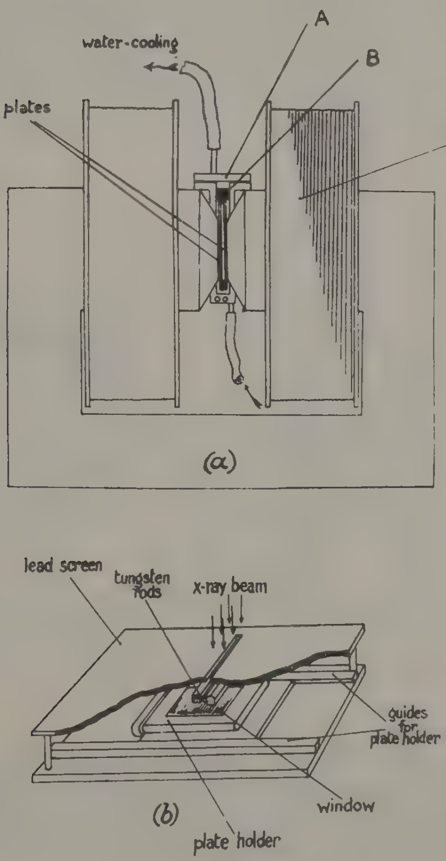
Magnetic Field. A field of 29,180 gauss, in a volume of $\pi \times (3{\cdot}5)^2 \times 0{\cdot}7$ cm.^{3}, was maintained by an electro-magnet, the necessary power being 600 watts, at a current of 8·5 amps. The field was constant during the exposure to within 300 gauss, apart from "edge effects", which were only appreciable in the region extending 4 mm. from the edges of the pole-pieces. The variation of the field with the current is shown in Table I., and it will be seen that it is very small for currents between 8 and 9 amps. During the exposure, the changes in current never exceeded 0·25 amps.

TABLE I.

Variation of field-strength with magnetizing current.

I (amps.)	8·0	8·2	8·4	8·5	8·6	8·8	9·0
H (gauss)	28,900	29,020	29,140	29,200	29,250	29,340	29,440

Fig. 1.


A
B
water-cooling
plates
coils
(a)
lead screen
tungsten rods
X-ray beam
guides for plate holder
window
(b)
plate holder

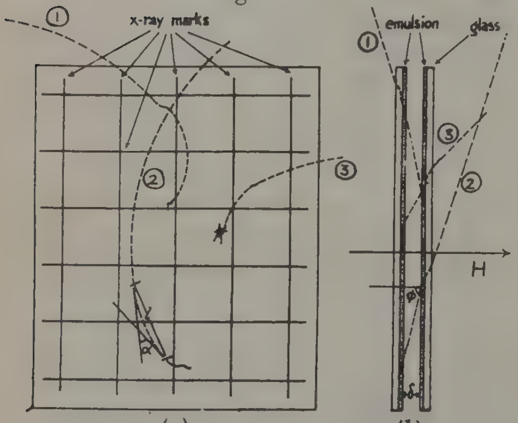
Experimental arrangements : (*a*) during the exposure,
(*b*) X-ray marking device.

The field was determined by comparison with a known field produced by a permanent magnet, kindly made available by Dr. S. Rosenblum and Mr. Tsai of Bellevue Laboratory in Paris, using a fluxmeter of high precision. A cooling system maintained the plates below 20° C.

Correlation of Pairs of Tracks. A precise knowledge of the geometrical relationship between the two plates during exposure was required in order to correlate the pairs of tracks produced by each particle, and this was obtained by the following method :—

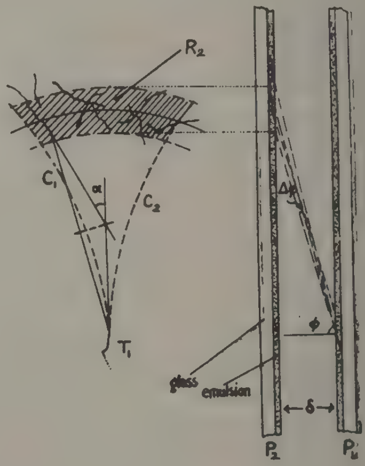
The holder containing the plates was exposed to an X-ray beam, passing perpendicular to the plane of the emulsion ; (see fig. 1(*b*)). Between the plates and the X-ray source was a screen of lead a few mm. thick, and provided with a narrow slit 8 cm. long and a few microns wide. With this arrangement it was possible to "print" a grid of lines on both the plates, as indicated in fig. 2, in such a way that corresponding points on the grids represented the same x, y, coordinates. The edges of the slit consisted of two cylinders of tungsten, 2 mm. in diameter and it was made sufficiently narrow by pressing a piece of thin paper between the two cylinders during assembly. The black lines produced in the processed plate by X-ray exposure had a width of 20μ, or even less.

Fig. 2.



(*a*) (*b*)

Fig. 3.



The rectangular grid of lines on the two plates greatly facilitated the correlation of the tracks. The two plates of each pair were first scanned, and maps were constructed showing the track of all particles crossing the surface of an emulsion, and of those which stopped in one of them. From the position, direction, and angle of dip, ϕ, of a track, T_1, which stopped in one emulsion, (see fig. 3) it was possible to determine a region, R_2, in the opposite plate, in which the continuation of the track should be found. If the density of the tracks in P_2 is low enough, only the true continuation of T_1 will have the correct ionization, angle of dip, and direction, consistent with it being the partner of track T_1. The probability that the track produced by another particle will, by chance, satisfy the same requirements is discussed in § 4, and is shown to be very small.

Processing of the Emulsions. Great care was taken in processing the plates in order to reduce the distortions of the emulsion to a minimum. The following procedures have been found satisfactory :

Chemicals	Temperature	Time		
		50 μ	100 μ	200 μ
For Kodak NT4.				
Developer (1 part D19b 2 ,, H_2O.)	18° C.	30m.	40m.	60m.
Hardening Stopping Bath	,,	10m.	15m.	20m.
Fixer (Hypo 30 per cent Solution in weight)	,,	3/2 of clearing time.		
Washing in water	,,	3h.	4h.	5h.
For Ilford G5.				
Developer	3° C.	—	—	35m.
16 cc. Azol Dev. 10 cc. 1 per cent K Br. Sol. 800 cc. H_2O.	18° C.	—	—	200m.
Hardening Stopping Bath Fixer	As for Kodak NT4.			

§ 3. METHOD OF CALCULATION OF MASS.

According to Camerini and Lattes (unpublished), the range-energy relationship for protons, in a photographic emulsion, can be represented by a simple power law for values of the range greater than 200μ :

$$E=KR^n,$$

where E is the energy of a proton of range, R. Further, the ranges R(v) and $r(v)$ of particles having the same initial velocity, v, but with masses M and m, and charges Ze and ze, respectively, are related by the equation

$$R(v)=r(v) \,.\, M/m(z/Z)^2-C,$$

C being constant *. Since C is negligible compared with the ranges of particles considered in the present experiment, it may be neglected. It follows that the relation between the energy ϵ and the range, r, of a particle of mass m, and charge ze, is given by

$$\epsilon=K(m/M)^{1-n}z^{2n}r^n, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

M being the mass of the proton, and e the elementary electronic charge.

The equation (3) is the explicit form of the equation (2). Solving the simultaneous equations (1) and (3), we obtain

$$\frac{m}{Z^\alpha}=A\left(\frac{H\rho}{\sin\phi}\right)^{\frac{2}{2-n}}\frac{n}{r^{2-n}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

---

* (The constant C takes account of the capture and loss of electrons in the last part of the range, a process which affects differently particles of different charge. See Bethe and Livingston (1937).)

where $\alpha = \frac{2(1-n)}{2-n}$ and $A = \left\{\frac{e^2}{c^2} M^{1-n}/2K\right\}^{\frac{1}{2-n}}$

For ranges shorter than 200μ, the range-energy relation given above is no longer valid, and the function can therefore only be evaluated numerically from the experimental points. If the logarithm of the residual range of a particle is plotted against the logarithm of its measured radius of curvature, it follows, from the equation (4), that the particles with equal masses should be represented by points lying on straight lines, of which the slope, ψ, is given by

$$\tan \psi = n/2.$$

There will be a departure of the observations from this linear relationship, however, in the case of particles of short range.

§4. SOURCES OF ERRORS.

The most important errors in the experiment are due to: (i) distortion of gelatine; (ii) scattering in the air-gap; (iii) scattering in the surface layers of the emulsion; and (iv) spurious coincidences.

(i) *Distortion of gelatine.*

Serious distortions of the emulsion extending over large areas, which can change the coordinates of a point by more than 10μ, have been found to be confined to regions near the edges of the plate. The affected area extends inwards from the edges of the plate for 2 mm. in the case of emulsions 50μ thick; and for 1 cm. for 200μ emulsions. On the other hand, local distortions, which may considerably alter the direction of a track, sometimes occur if special care is not taken in processing. Local distortions are very often accompanied by differences in the general background of grains; and they may sometimes be detected, with the naked eye, in the form of manifest imperfections in the smooth surface of the emulsion.

(ii) *Scattering in the air-gap.*

In crossing the gap between the plates, the particle may be scattered by the gas filling it. The influence of such effects on the measurement of the curvature of the tracks of particles in magnetic deflection experiments, has been discussed by Bethe (1948). His calculations, when applied to our particular experimental conditions, lead to the result given in Table II., which shows the values of $\frac{\rho_H}{\rho_S}$, the ratio of the radius of curvature of a track due to the magnetic field, to the average apparent radius of curvature due to scattering. $\frac{\rho_H}{\rho_S} = \frac{\delta_S}{\delta_H}$, where δ_S and δ_H represent the angular deviations due to scattering and the magnetic field, respectively.

(ii) *Scattering in the surface layers of the emulsion.*

The determination of the direction of motion of a particle, at the exit from one emulsion and at entry into the other, is subject to errors as a result both of scattering by the atoms of the emulsion, and of the displacement of the individual grains in the track from the true line of the trajectory. The second of these effects has been studied by Goldschmidt *et al.* (1948), in experiments with C2 Ilford plates, and has been found to be less than three per cent of the real scattering for tracks of range less than 2000 μ. The error from "spurious scattering" can therefore be regarded as negligible.

The errors due to scattering in the surface layers in the emulsion depend on the method of defining the direction of the track. If we choose to define this as the direction of the chord in the last few microns as the track approaches the surface of the emulsion, the average error can be evaluated :

From the formulæ given by Rossi and Greisen (1941), and using their notation we easily obtain for the angle ξ between chord and trajectory

$$\langle \xi^2 \rangle = 4/3(t/w^2),$$

TABLE II.

Values of $\rho_H/\rho_S = \delta\,\delta_S/\delta_H$; H=29,200 ; l=1 cm.

Residual range	Mass of Particle		
	200 m_e	300 m_e	2000 m_e
μ	(per cent)	(per cent)	(per cent)
100	2·7	2·9	4·3
200	2·3	2·55	4·0
400	2·0	2·15	3·4
600	1·85	2·03	3·0
800	1·7	1·9	2·9
1000	1·0	1·8	2·7

The calculations have been made for a particle of charge e, and masses $\sim 200\,m_e$, $300\,m_e$ and $2000\,m_e$ for a path length in the air-gap of 1 cm. The values of ρ_H/ρ_S, when multiplied by $2/(2-n)$ (about 1·4 in the present instance) give the error in the mass due to the scattering in the air.

where t is the length of the chord in radiation units, and w is a function of the momentum and velocity. On the other hand, for the total angular deviation θ of a track from its original direction, after a path, t, we have

$$\langle \theta^2 \rangle = 2t/w^2 = 1{\cdot}5 \langle \xi^2 \rangle .$$

It follows that $\langle \xi^2 \rangle$ is easily evaluated from experimental values of $\langle \theta^2 \rangle$ (Goldschmidt *et al.* 1948).

A simple and more direct method of estimating the importance of all the above sources of error is to make an exposure without a magnetic field. The measured "deflections" are then due to scattering or distortion. An examination of a plate containing a total of 50 tracks (10 mesons and 40 protons) has shown that the errors in the angular measurements are distributed in a gaussian curve, the half-width of which is of the order of 1·4°.

(iv) *Spurious coincidences.*

The number of "spurious coincidences"—the number of cases in which two unrelated tracks are attributed to the same particle—depends on the criterion adopted for the recognition of true pairs—and the method employed has been described in §2. Consider a track T_1 in P_1, and let its angle of "dip" be $\phi_1 \pm \Delta\phi$, its residual range, r_1, and its direction α_1. The region on P_2 in which a search for the partner track will be made, is the area R_2, shown in fig. 3; this area is limited by two circles of radius $\delta \tan(\phi_1 - \Delta\phi)$ and $\delta \tan(\phi_1 + \Delta\phi)$ on the one hand; and by the circular paths C_1 and C_2, corresponding to the maximum deflection expected for the particles with mass greater than 100 m.e., on the other.

Suppose now that the tracks in P_2 are isotropically distributed. This assumption is only approximately correct, but it can hardly lead to serious errors. Suppose, further, that the tracks occur with a frequency per unit area equal to ν_2. The number of "fortuitous coincidences" will then be given by the number of tracks contained in R_2, crossing the surface of the emulsion, with the correct angle of dip, $\phi \pm \Delta\phi$, and correct direction, $\alpha_2 \pm \Delta\alpha$. If $\Delta\phi$ is small, the required value is given to a sufficient approximation by

$$n = \frac{2\nu}{\pi} \cdot \frac{\Delta\alpha(\Delta\phi)^2 . \delta^3}{\rho_{\min}} \cdot \frac{\tan^3\phi}{\cos\phi}.$$

For N tracks, the total number of "coincidences", C, is,

$$C = \sum_1^N n.$$

If the N tracks are isotropically distributed within the angles ϕ_1 and ϕ_2, we then obtain

$$C = \frac{2N\nu . \Delta\alpha . (\Delta\phi)^2 \delta^3}{\pi\rho_{\min}} \cdot \left[\tfrac{1}{3}\tan^3\phi - \tan\phi + \phi\right]_{\varphi_1}^{\varphi_2}.$$

In the conditions of our experiment:

$$\Delta\alpha = 1° = 0{\cdot}017 \text{ radians},$$
$$\Delta\phi = 2° = 0{\cdot}034 \text{ radians},$$
$$\delta = 0{\cdot}3 \text{ cm.},$$
$$\rho_{\min} = 2 \text{ cm.}$$

Considering only particles having a path, l, in air between 1 and 4 cm., we find, from Table III., that $C \sim 10^{-4}\nu N$.

TABLE III.

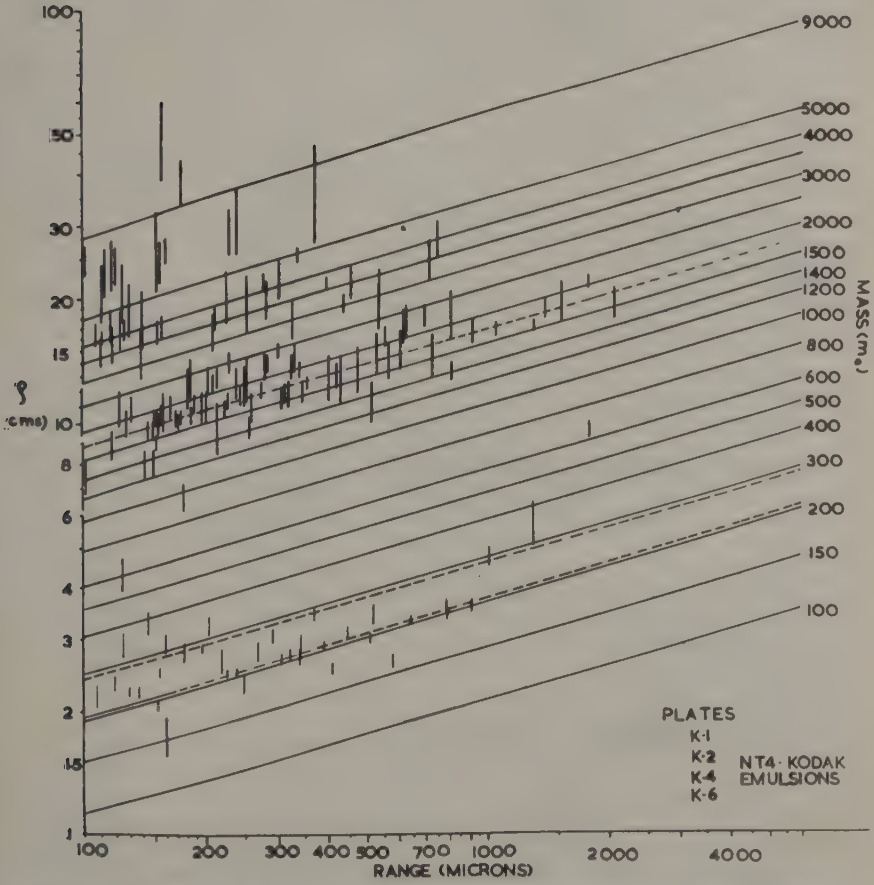
Values of $\Phi = [\frac{1}{3}\tan^3\phi - \tan\phi + \phi]$

l (cm.)	1·0	1·5	2·0	3·0	4·0
ϕ (rad)	1·27	1·37	1·42	1·47	1·50
$\Phi(\phi)$	10·4	35·2	92·7	324·5	788·0

If $\nu < 100$ cm.$^{-2}$ the number of fortuitous coincidences will be less than 1 per cent.

In this calculation, it has been assumed that the correlation of the events has been made only on the basis of geometrical considerations, whereas in actual fact, the observation of a close similarity in the grain-density of the two tracks of a pair has always been considered an essential additional criterion. We have assumed, however, an isotropic distribution of the tracks which (see fig. 7) does not, in fact, exist. Approximate estimates suggest that the two effects compensate one another, and it is reasonable to suppose that the calculations give a satisfactory indication of the order of magnitude of the number of spurious coincidences.

Fig. 4.

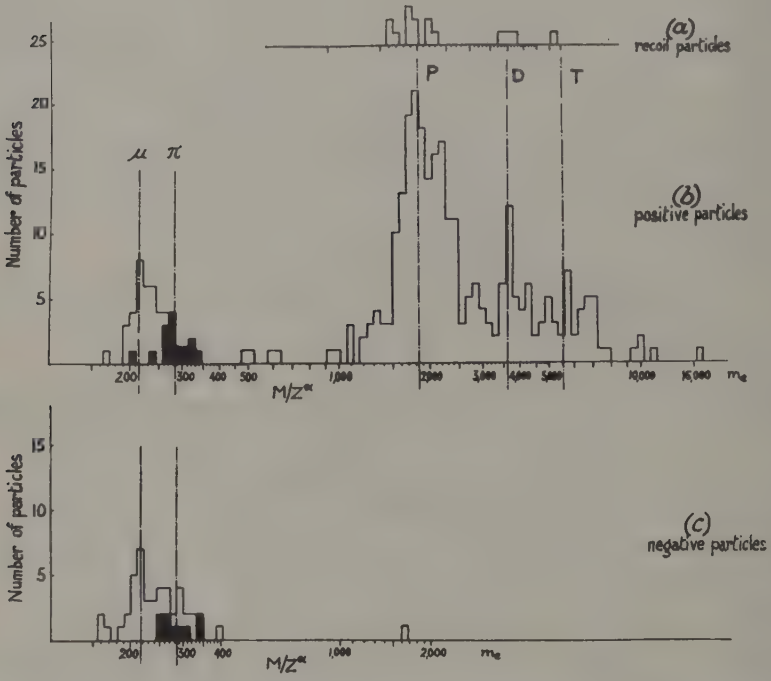


Logarithm of radius of curvature ρ (cm.) versus logarithm of residual range r. The lines represent the function with slight corrections to allow for deviations of the range-energy relation from the power law for short ranges.

§ 5. RESULTS.

In fig. 4 the results of measurements on the tracks of 150 particles, observed in Kodak NT4 emulsions, have been plotted. The probable error in the measurement of the radius of curvature, calculated in each case as discussed in § 4 is indicated for each point. For the sake of clarity, the error in range, which is estimated to be smaller than 5 per cent in every case, has not been shown. A similar graph has been constructed for the other 210 measurements, made with Ilford G5 plates, for which the range-energy relation is slightly different, and the results are essentially similar. It will be seen that the results clearly show the division of the particles into groups corresponding to mesons, protons and heavier particles, the most prominent group being that due to protons. The linear distribution, anticipated from equation (4), for particles of a given mass and of range greater than 200 μ, is clearly displayed.

Fig. 5.



Mass distribution of 360 recorded particles. The dotted lines correspond to mass values of μ-mesons (215 m_e), π-mesons (285 m_e), protons, deuterons, tritons. It follows from equation (4) that a particle of charge Ze appears (using this method) to have a mass of m/Z^α, α being 0·59 for Kodak NT4, and 0·56 for Ilford G5. $\pi+$ and $\sigma\_$ mesons are indicated in black.

The results of measurements of the mass of 359 particles are shown graphically in fig. 5, and details of the measurements on the group of particles with masses less than 400 m_e, including the secondary phenomena observed at the end of their range, are given in Table IV.

TABLE IV.

Analysis of Mesons—Mass Measurements.
Plates Sensitive to Minimum Ionization Tracks.

The mesons are classified phenomenologically, according to the secondary effects observed at the end of their range, into π, σ, ρ (Powell, 1949). A μ-meson is one which is *observed* to originate by the decay of a π-particle in the emulsion.

No.	Sign	α (Deg.)	ρ (cm.)	Residual Range	Mass	Probable error in Mass	Observations
K.O. 12	—	13·0	1·94	105	220	20	ρ electron decay
K.1. 11	+	17·2	2·18	108	240	30	ρ electron decay
21	—	27·5	2·55	415	178	20	ρ
29	+	29·5	2·94	308	215	10	ρ
26	+	23·0	3·63	921	215	18	ρ
76	+	19·0	2·24	130	235	15	ρ electron decay
90	+	21·0	2·80	178	285	35	π–μ decay
110	—	24·1	2·22	137	230	15	ρ (slow electron)
158	—	30·2	2·71	344	210	18	ρ electron decay
164	+	37·5	2·56	258	220	10	μ
72	—	42·0	2·96	126	350	50	σ (5 branch star)
55	—	26·7	2·76	327	216	10	ρ (blob)?
125	—	61·5	3·63	650	215	7	ρ (slow electron)
142	+	51·5	3·1	395	206	10	μ (electron decay)
39	—	35·5	3·84	249	390	30	ρ (blob)
K.2. 24	—	22·5	3·09	295	270	10	σ (3 branch star)
89	—	25·0	2·54	585	158	25	ρ
75	+	21·0	2·3	57	240	30	π
I.7. 2	+	15·8	3·55	388	280	20	π
6	+	8·0	3·7	286	330	30	π
96	+	33·0	2·5	305	190	30	ρ electron decay
118	+	15·0	2·7	315	120	18	ρ electron decay
133	+	19·0	1·9	220	155	20	ρ electron decay
141	—	16·0	3·94	690	250	20	σ (1 branch star)
156	—	9·3	3·8	233	320	30	ρ slow electron
186	+	29·0	2·7	310	212	20	μ electron decay
10	+	10·0	2·85	335	220	20	μ electron decay
35	—	23·7	1·9	145	205	20	ρ
54	—	44·4	1·85	101	220	7	ρ (blob)?
121	—	10·0	3·33	357	265	30	ρ (blob)
134	+	12·0	2·5	124	310	30	π
190	—	13·0	2·35	235	200	40	ρ (blob)?
191	+	11·0	2·7	206	250	20	ρ
173	+	12·0	2·94	335	225	15	ρ electron decay
176	+	8·0	2·88	290	235	20	ρ electron decay
177	—	19·6	2·88	355	220	12	ρ electron decay
I.8. 64	—	31·6	2·9	298	235	10	ρ slow electron
76	+	13·4	2·95	165	320	25	π
81	+	13·9	2·58	270	210	25	μ electron decay
85	—	13·5	3·55	430	265	25	σ (2 br. star)
93	+	24·4	4·44	1530	190	15	ρ electron decay
91	+	10·0	2·6	241	230	20	ρ electron decay
99	—	12·5	1·9	200	165	25	ρ

TABLE IV. (*Cont.*)

No.	Sign	α (Deg.)	ρ (cm.)	Residual Range	Mass	Probable error in Mass	Observations
154	−	22·5	2·78	396	195	10	ρ (blob? and slow electron)
164	+	14·0	2·14	160	170	25	ρ electron decay
174	+	16·5	3·0	370	225	15	ρ
171	+	14·0	2·05	510	250	30	μ
140	−	19·0	1·7	120	180	50	ρ
226	−	12·0	2·55	140	292	50	σ (1 heavy trk.)
I.9. 29	−	17·0	2·75	202	260	20	ρ
46	+	15·0	4·45	714	280	20	π
47	+	17·5	2·6	200	245	20	μ (incomplete)
52	+	14·4	3·0	400	215	5	ρ electron decay
65	+	25·5	2·45	295	192	10	ρ electron decay
74	−	9·0	4·0	450	300	20	ρ
79	+	10·5	2·35	236	200	40	π
95	−	17·1	4·0	508	290	20	ρ slow electron
105	−	9·0	3·0	108	350	40	σ (1 heavy br.)
121	−	19·0	3·33	392	245	20	ρ (blob)?
125	−	19·8	4·0	484	295	30	ρ (blob)
3	−	14·5	3·28	496	218	25	ρ (blob)?
8	+	20·9	1·95	135	207	15	ρ electron decay
16	+	7·0	3·7	460	260	20	ρ electron decay
30	−	13·5	4·5	902	250	40	ρ electron decay
95	+	9·0	4·75	1700	200	20	ρ electron decay
84	−	9·0	3·2	227	310	30	σ (4 branch star)

Plates not Sensitive to Minimum Ionization Tracks.

No.	Sign	α (Deg.)	ρ (cm.)	Residual Range	Mass	Probable error in Mass	Observations
K.4. 26	−	14·0	2·34	120	262	25	σ (4 branch star)
63	−	13·7	1·75	160	160	30	ρ
71	−	12·0	2·33	250	195	20	ρ
30	−	24·5	2·99	512	267	15	ρ (slow electron)
31	−	16·5	2·81	267	240	25	ρ
38	+	91·6	2·5	228	220	10	ρ
41	−	22·5	2·3	150	230	15	ρ
44	+	26·0	3·65	860	225	15	ρ
91	−	13·4	3·48	800	208	15	ρ
28	−	21·0	3·48	374	282	15	ρ
107	−	10·8	2·66	220	245	30	σ (3 branch star)
47	+	20·0	2·92	346	230	20	ρ
111	+	9·5	5·7	1300	340	80	π
100	+	14·9	4·72	1030	292	20	π
K.6. 68	−	21·5	3·26	205	330	30	ρ
70	+	17·0	2·88	197	283	7	π
114	−	11·0	2·05	85	250	50	ρ
92	+	31·0	1·75	45	258	20	π
97	+	33·0	2·17	91	262	13	π
19	+	17·0	2·5	160	260	15	π
26	−	28·5	3·1	450	228	15	ρ
45	−	30·0	2·67	153	200	17	ρ
53	+	12·5	3·45	523	235	16	μ
57	−	11·0	2·94	160	315	40	ρ
109	−	26·4	2·5	239	217	10	ρ

The observed distributions of the values of the mass of the particles, taken separately according to their sign, are shown in figs. 5(*b*) and (*c*).

The distribution in mass of the positively charged particles (fig. 5(*b*)) shows :

(*a*) A peak due to particles with masses in the interval from 160 to 250 m_e, the mean value being $217 \pm 4\ m_e$*. About 80 per cent of these particles decay with the emission of an electron, of which the track can be distinguished. Six of them are μ-mesons created by the decay of π-particles in the emulsion.

(*b*) A peak, just resolved from the first, and due to particles with masses between 240 and 340 m_e, the mean value being $281 \pm 10\ m_e$. All these particles decay at the end of their range with the emission of a μ-particle.

(*c*) A more prominent group, corresponding to the particles with a mean mass of 1870 m_e, $\pm 27\ m_e$. None of these particles give any indication of inter-acting with matter at the end of their range, nor do they produce any observed charged particles by spontaneous decay.

(*d*) A peak of lower intensity can be seen at about 3700 m_e, and an indication of a broad distribution due to particles of a greater mass. None of these particles give any indication of inter-acting with matter, nor of undergoing spontaneous decay.

The distribution of values of the mass of negative particles (fig. 5(*c*)) shows :

(*e*) A peak between 160 and 250 m_e, and a broader distribution extending up to 400 m_e. Of the particles with masses between 160 and 250 m_e, only four show electron decay ; and five others give rise to slow electrons originating at the end of their range. Of the particles between 250 and 400 m_e, 9 produce nuclear disintegrations, and the others stop without leading to the production of any secondary charged particles for which the evidence is unambiguous. In ten cases there is an unusual thickening at the end of the track, see Pl. II.(*b*), due to an assembly of two or three grains which might be attributed to a secondary particle of very short range ; but in all cases except three it has been impossible to decide whether this " blob " was produced by a slow electron or by a heavy particle. In the three cases the " blob " was definitely too thick to have been produced by an electron.

(*f*) In fifteen cases, a particle which crosses the air-gap and is brought to rest in one emulsion, originates in the emulsion of the other plate. The complete track therefore has two " ends ". The direction of motion of the particle, in events of this type, can be determined by the increased grain-density in the track as the particle slows down, and the more numerous deviations due to Coulomb scattering towards the end of the range. The particles of this type are all positively charged and the distribution in the observed values of their mass is shown in fig. 5(*a*).

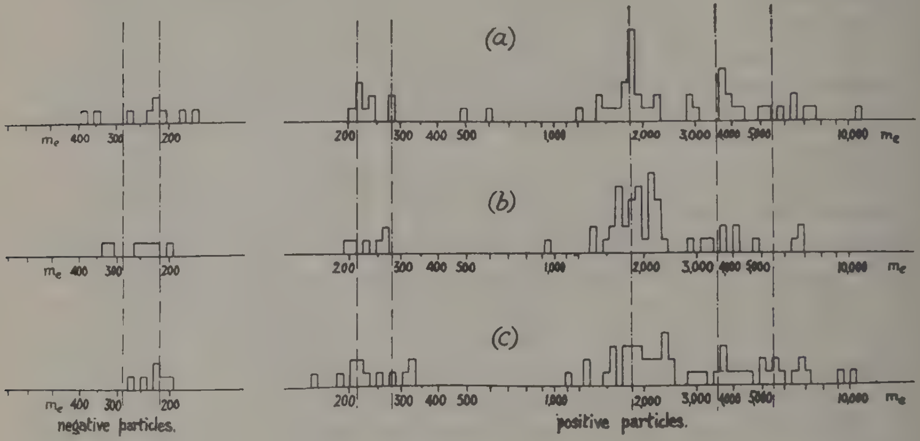
---

* The errors, σ, are calculated from the residuals, v_i, according to the formula

$$\sigma = \sqrt{\frac{\Sigma v_i^2}{n(n-1)}}.$$

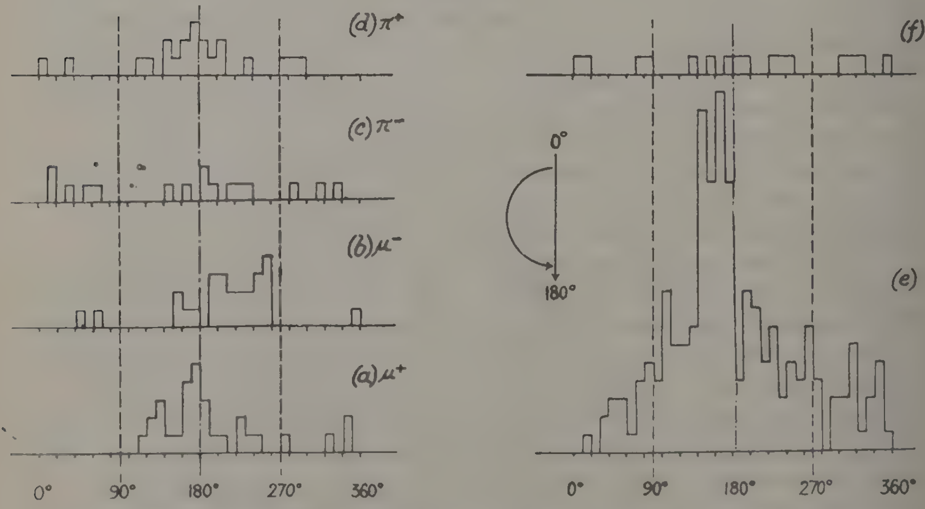
(*g*) The angular distribution, relative to the vertical, of the directions of motion of all particles on which measurements have been made, at their points of entry into the first emulsion through which they pass, are shown in fig. 7. In this figure, the particles have been grouped together according to their masses and the signs of their charge. In

Fig. 6.



Effects of distortion of the gelatine on the precision of the measurements.
(*a*) 61 tracks from 2 pairs of 50 μ thick emulsions (25 days exposure).
(*b*) 56 tracks from 1 pair of 100 μ thick emulsions (15 days exposure).
(*c*) 69 tracks from 1 pair of 200 μ thick emulsions (10 days exposure).

Fig. 7.



Angular distribution of: (*a*) μ^+; (*b*) μ^-; (*c*) π^-; (*d*) π^+; (*e*) recoil particles; (*f*) protons, deuterons and heavier particles. This histogram also contains some tracks which were not in a suitable position for being measured, although their identity could be established with certainty.

most cases there is a tendency for the directions of motion to be grouped round the vertical, but to be deflected from it in a sense depending on the sign of the charge.

(*h*) There is a gradual reduction in precision as the thickness of the emulsion employed is increased—see fig. 6.

§6. INTERPRETATION OF RESULTS.

From the evidence given in sections (*a*) to (*g*) above it is concluded that:

(α) The positive particles corresponding to group (*a*) are μ^+-mesons created by the decay of π^+-particles. In those cases in which the μ^+-particle has an energy greater than 4·2 MeV., we must assume that the parent π^+-particle decayed whilst in motion. The average value of the mass of these particles is $217 \pm 4\ m_e$. We therefore identify them with the mesons of the penetrating component of the cosmic radiation, of which the mass, as determined at Berkeley, is $215 \pm 4\ m_e$. 80 per cent of these particles decay with the emission of an electron. In view of the fact that "thin" tracks inclined steeply to the plane of the emulsion are difficult to detect, and since there is an isotropic distribution of the directions of motion of the decay electrons, we may assume that we fail to detect 20 per cent of the electron tracks, of which the specific ionization has the minimum value, in the conditions of the experiment; but that, in fact, all the μ^+-mesons decay with the emission of an electron.

(β) The particles described in group (*b*) are obviously π^+-particles. From the observed value of the mass, $281 \pm 7\ m_e$, they can be identified with those produced artificially.

(γ) The negative particles described in group (*e*) are made up of μ^-- and π^--particles. Nine of them produce observed disintegrations, and of these the average value of the mass is $288 \pm 13\ m_e$. They can therefore be identified as π^--particles. The average mass of the other 38 negative particles, only four of which appear to decay with the emission of an electron, is $236 \pm 8\ m_e$. It is reasonable to attribute this value, intermediate between the masses of the π and μ particles, to the presence among the μ^--mesons, of some π^--particles which produce no visible disintegrations. The observed value of the average mass suggests that 28 ± 4 of these mesons are μ^--particles, and 10 ± 4 are π^--particles.

(δ) The group with apparent masses between 1100 and 2400 m_e., which contains 48 per cent of all tracks measured, is formed by positive particles which stop in the emulsion without showing any inter-action with matter, nor any evidence of spontaneous decay. Their average mass is $1870 \pm 27\ m_e$, and they may be identified with confidence as protons.

(ϵ) The group between 2400 and 4400 m_e, of smaller intensity than the protons and less clearly resolved, is also made up of positive particles which produce no secondary effects; they are therefore regarded as deuterons.

No accurate discrimination of particles of different types is possible in the region of the spectrum beyond 5500 m_e. This is due not only to the small deviations suffered by heavy particles, but also to the fact that their charge, which may be different from $|e|$, is unknown. We therefore attribute this group to a mixture of tritons and α-particles, which cannot be resolved in the conditions of our experiment.

(ζ) Of the seven particles which appear to have an intermediate value of the mass, six are positively charged, and one appears to be negatively charged. We must expect that a certain proportion of the different particles which we record will be scattered through an angle of a few degrees or more, in passing through the gas between the plates, as the result of elastic collisions with oxygen or nitrogen nuclei. Such a deviation in the trajectory will usually lead us to reject the particle from among those which give tracks acceptable for measurement, for the directions of motion—at the points of exit from and entry into the two emulsions—will not in general lie on a circle. In a small proportion of such cases however, the deviation will be of such a character that the tracks are still acceptable. This will be so, for example, if the nuclear collision occurs near the mid-point of the trajectory of the particle in the air-gap. The observed deviation will then be less or greater than that due to the magnetic deflection alone, and the calculated values of the mass will be correspondingly greater or less than the true value.

Other processes, such as inelastic scattering of protons in the emulsion, will tend to lead to an over-estimate of the masses of the particles. In addition, as has been shown in §4, we must expect to observe a certain number of fortuitous coincidences, and their number in the particular conditions of the experiment should be 3 or 4. Because of these and other possible processes, it appears reasonable to attribute the observation of particles with intermediate values of the mass to scattering in the gas, or in the surface layers of the emulsion, or to other uncontrolled sources of error. The results are therefore regarded as giving no support for the view that other particles exist which are sufficiently stable to be brought to rest in solid materials, and which have masses in the interval between π-mesons and protons; and as proving that if such particles exist they either occur in matter bombarded by cosmic radiation at 11,000 feet, with a frequency less than 3 per cent of that of π and μ particles and less than 1 per cent of protons; or, alternatively, they have a lifetime shorter than 10^{-10} sec.

(η) Of the particles which originate in the emulsion (see fig. 5(a)), 11 have masses equal, within the limits of error, to that of protons. Three of the others may be identified as deuterons and one is of greater mass—probably an α-particle. The majority of the particles may be attributed to knock-on protons produced by the impact of neutrons with hydrogen nuclei; and others to nuclear disintegrations produced by neutrons of moderate energy, in which a single charged particle is emitted.

(ϑ) The distribution in the directions of motion of the μ^+ and μ^--mesons—see fig. 7 (a) and (b)—shows that most of these particles move downwards at a small inclination to the vertical—see (Camerini *et al.* 1948).

There is some indication that the maxima in the two distributions deviate from the vertical in opposite senses according to the sign of the charge. This may be accounted for if a considerable proportion of the particles are generated in the atmosphere or other material above the apparatus. They will then travel in the magnetic field for a distance of several centimetres before entering the emulsion, and will suffer a corresponding deflection.

The statistical weight of the angular distributions of the π^+ and π^--particles is too small for any conclusions to be drawn, but the observations are consistent with the assumption that the majority of particles are locally generated in the matter in the immediate neighbourhood of the plates, and that the directions of motion are distributed approximately isotropically.

TABLE V.

Type of particle	Number of particles	Number with			Percentage of all particles
		electron decay	slow electron	‘ blob ’	
μ^+	30	19(24)*	—	—	8·5±1·5
μ^-	28±4	4(18)*	6(18)*	6(18)*?	8±2·5
π^+	14	—	—	—	4·0±1·0
π^-	19±4	—	—	3	5·4±2·0
H^1	168	—	—	—	48±4
H^2	55	—	—	—	16±2
H^3,α etc.	39	—	—	—	11±2

* These values are deduced from observations on electron-sensitive emulsions only. The figure in brackets gives the corresponding total numbers of particles observed in these plates.

The distributions for the protons, fig. 7(*e*), can be regarded as being made of two groups, one of which is due to particles moving in directions inclined to the vertical at less than 40°. From the displacement of the of the peak from the vertical, 10°, we must assume that most of the particles have travelled at least 5 cm. in the magnetic field. The other group, containing about 60 per cent of the total, is made up of particles of which the directions of motion are more widely dispersed, and they may be attributed to protons which originate in nuclear explosions produced in the matter of the apparatus.

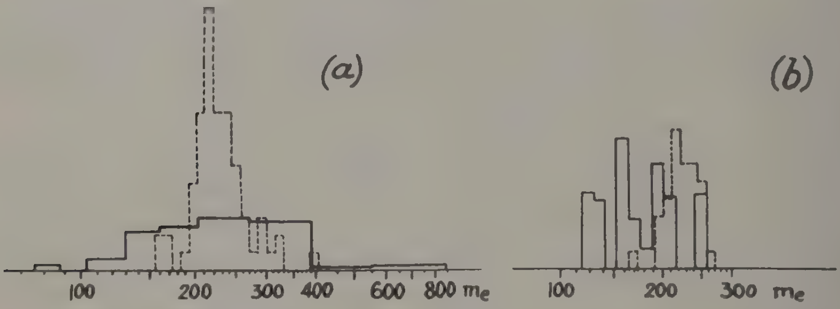
The results of the present experiments are summarized in Table V., which shows details of the proportions of the particles of different types amongst all those on which measurements have been made.

§6. COMPARISON WITH OTHER METHODS.

A comparison of the present results on the masses of the particles with those obtained in other experiments employing photographic emulsions is shown in fig. 8. The figure gives the mass distributions obtained by Goldschmidt *et al.* from scattering measurements, and by Lattes, Occhialini and Powell from "grain-counts". In each case it can be seen that the "width" of the distribution, due to particles of a single type, is much greater than that obtained by the magnetic deflection experiments.

Experiments similar in principle to those described in the present paper have been carried out by Alichanian and Alichanov and their collaborators (1948). Using Geiger–Muller counters, they have measured the deflection, produced by a magnetic field, of particles brought to rest in a certain thickness of lead. The mass distribution thus obtained shows, (*a*) a prominent peak corresponding to the protons; (*b*) groups of particles of greater mass than protons, but not individually resolved; and (*c*) a continuous distribution towards lighter masses, which fills completely the interval between the groups due to protons and μ-mesons. The Soviet experimenters have interpreted their results as indicating the existence of many types of particles with intermediate masses which can carry both positive and negative charges. Further, a preliminary report has been given by Brode (1949) of experiments of the same type, but capable of greater precision, in which the deflections are measured by observations with Wilson chambers. In this case, it has been suggested, tentatively, that the observations may correspond to the existence of positive particles of mass about 1000 m_e. In the case of both these experiments the particles are arrested in layers of lead, and it is therefore not possible, in general, to observe any secondary processes occurring at the end of the range of the particles.

Fig. 8.



Comparison with other methods. (*a*) The mass distribution of all the ρ-mesons, (*i. e.* $\mu+$, μ^- and that fraction of π^- which does not produce stars) obtained by magnetic deflection (dotted line), and that obtained by scattering measurements (full line) (Goldschmidt *et al.* 1948). (*b*) The mass distribution of the $\mu+$ obtained by magnetic deflection (dotted line), and that obtained by grain-counting (Lattes *et al.* 1948).

The conclusions to which we have been led by the results of the present experiments are therefore in contradiction with those obtained by other methods. The seven observed particles with intermediate values of the mass have been attributed to uncontrolled sources or error. It appears to be of great significance that all but one of the particles are positively charged, and that they show no evidence of spontaneous decay at the end of their range. A similar feature is characteristic of the results described by Brode, in which all the observed particles of intermediate mass are positively charged. In view of the great variety of physical processes which can occur in apparatus of the type described by Alichanian and his colleagues, in addition to those which alone are assumed to be operative in the interpretation of the results, it may be suggested that the existence of other types of mesons, in addition to the π- and μ-particles, should be accepted with very great reserve; and that the present experiments prove that if they exist, they are certainly very much more rare than previous experiments have been taken to indicate.

ACKNOWLEDGMENTS.

These experiments were originally undertaken in collaboration with Dr. S. Rosenblum, who installed the electro-magnet at the Jungfraujoch High-altitude Research Station, and who contributed both to the design of the apparatus and in making the exposures. I have great pleasure in acknowledging this substantial assistance without which it would have been impossible to complete the experiments.

I am greatly indebted to Professor C. F. Powell for his continuous help and encouragement both during the course of the work and in the discussion of the results. I also wish to thank Dr. H. Heitler for his help in the construction of some of the more delicate parts of the apparatus; Messrs. Dick and Roberts who assisted with the exposures and in the installation of the electromagnet; to Mrs. Andrews, Mrs. Reid and Miss Merritt for valuable assistance in the scanning of the plates and Professor Von Muralt and Mr. Widerkehr for kind hospitality. Finally, I wish to make grateful acknowledgment to Mr. Abele Giandolini of London, for financial assistance during the earlier part of the work.

REFERENCES.

ALICHANYAN, ALICHANOV and WEISSENBERG, 1948, *Journ. Exp. Theor. Phys., U.S.S.R.*, **18**, 301.

ALICHANYAN, ALICHANOV, MOROZOV and KHRIMIAN, 1948, *Acad. Sci. Rep.*, **61**, 35; and earlier references.

BETHE, 1946, *Phys. Rev.*, **70**, 821.

BETHE and LIVINGSTON, 1937, *Rev. Mod. Phys.*, **9**, 269.

BRODE, 1949a, *Rev. Mod. Phys.*, **21**, 37; 1949b, *Communication at Como Conference, September.*

CAMERINI, MUIRHEAD, POWELL and RITSON, 1948, *Nature*, **162**, 433.

GOLDSCHMIDT-CLERMONT, KING, MUIRHEAD and RITSON, 1948, *Proc. Phys. Soc.*, **61**, 138.

LATTES, OCCHIALINI and POWELL, 1948, *Proc. Phys. Soc.*, **61**, 173.

POWELL and ROSENBLUM, 1948, *Nature*, **161**, 473.

RETALLACK and BRODE, 1949, *Phys. Rev.*, **75**, 1716.

ROSSI and GREISEN, *Rev. Mod. Phys.*, 1941, **13**, 240.

PLATES I. & II. (*a*).

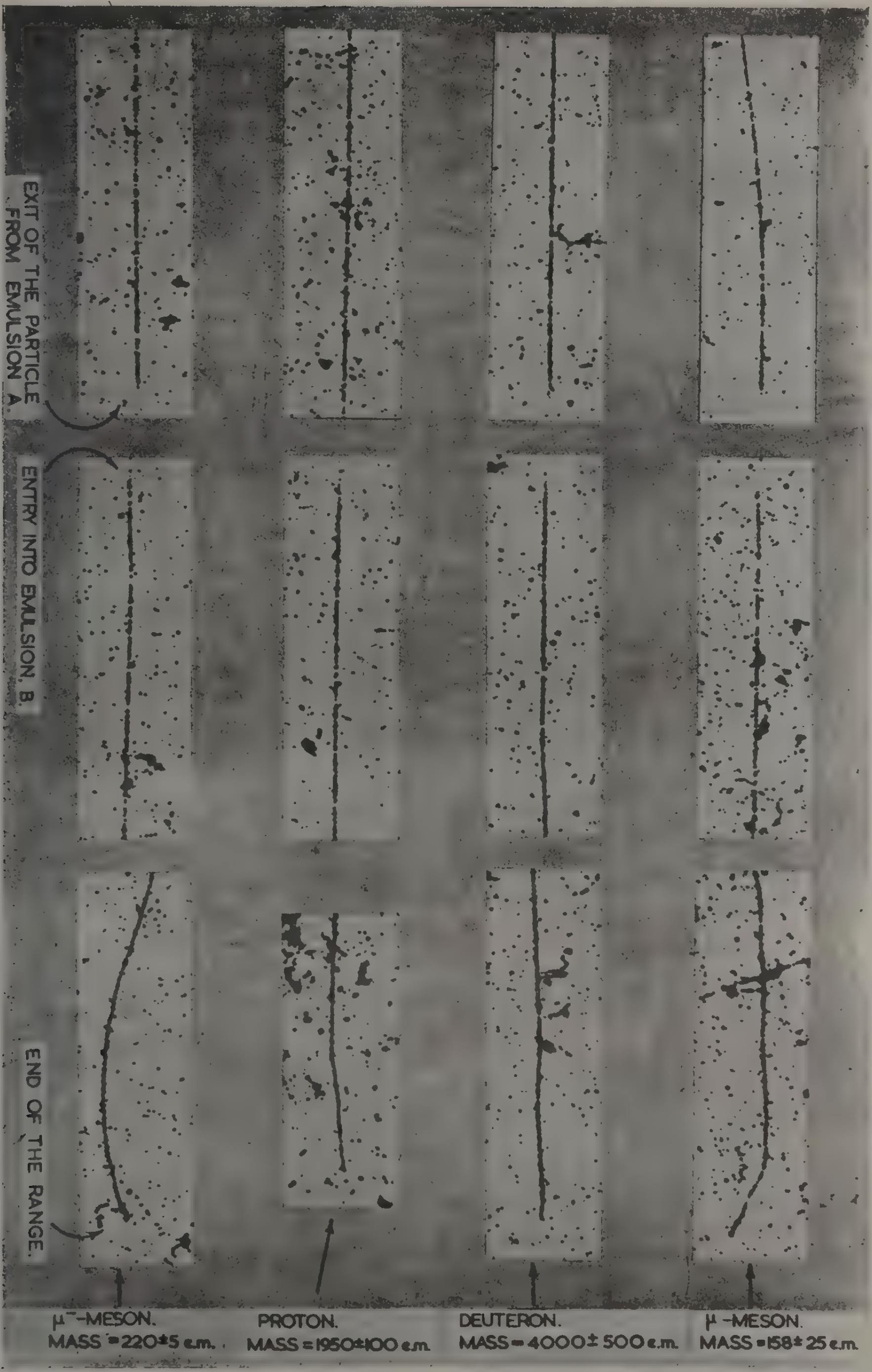
Examples of "partners".

Photographs of the tracks of particles at their exit from the first emulsion (first row), at the entry into the second emulsion (second row), and at the end of their range.

The great similarity between "partners", which is immediately apparent under the microscope, has been partially lost in the photograph.

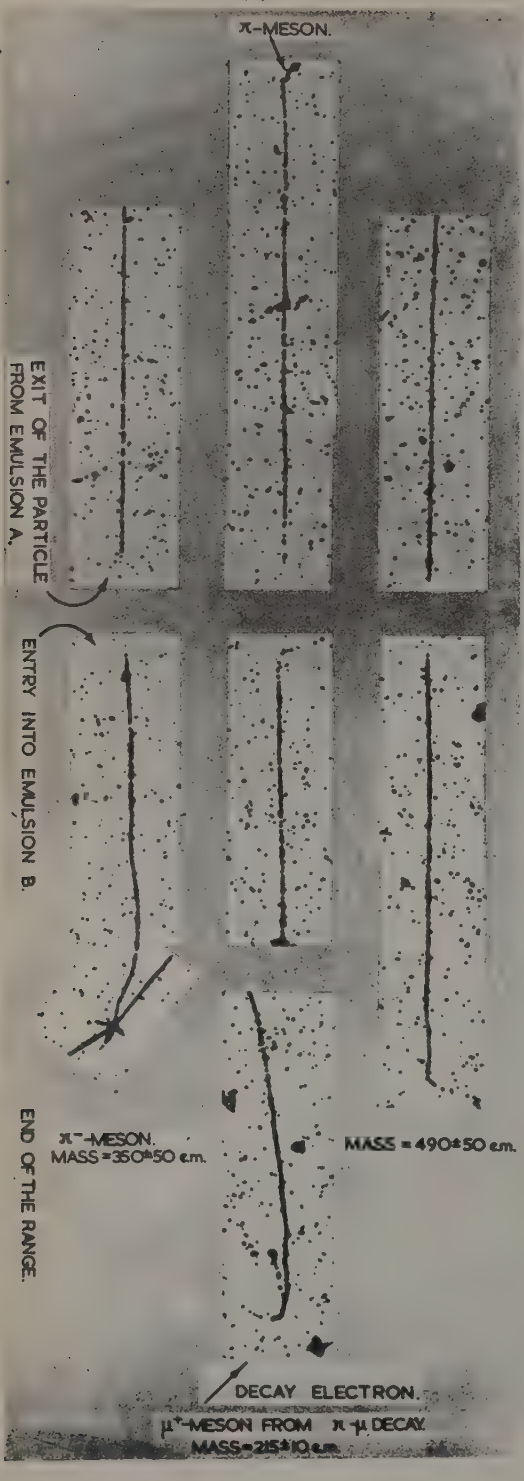
PLATE II. (*b*).

Negative meson. The arrow points to the "Blob" at the end of the range, which seems to be characteristic of negative mesons only.

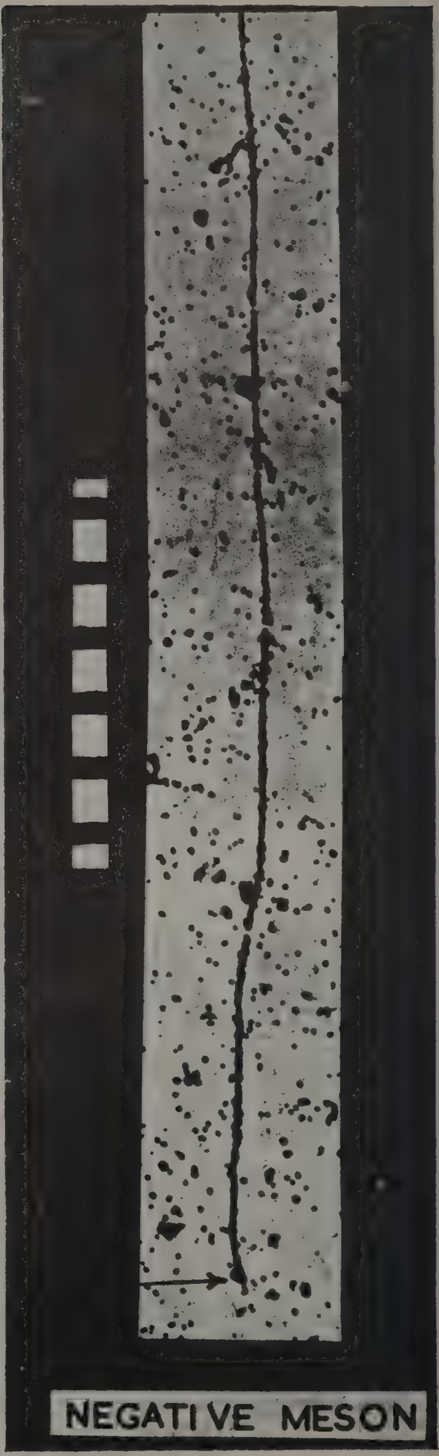

EXIT OF THE PARTICLE FROM EMULSION A.
ENTRY INTO EMULSION, B.
END OF THE RANGE.
μ^--MESON. MASS =220±5 e.m.
PROTON. MASS =1950±100 e.m.
DEUTERON. MASS = 4000 ± 500 e.m.
μ -MESON. MASS =158± 25 e.m.

To face page 106

(*a*)


π-MESON.
EXIT OF THE PARTICLE FROM EMULSION A.
ENTRY INTO EMULSION B.
END OF THE RANGE.
π⁻-MESON. MASS = 350±50 e.m.
MASS = 490±50 e.m.
DECAY ELECTRON.
μ⁺-MESON FROM π-μ DECAY. MASS = 215±10 e.m.

(*b*)


NEGATIVE MESON

VIII. *Notices of New Books and Periodicals received.*

Supersonic Flow Around Cones of Large Yaw. (Massachusetts Institute of Technology Technical Report No. 5.) [Pp. 125.] (Cambridge: Massachusetts.)

The properties of the flow about a yawing cone can be obtained as a power series in the yaw ϵ; earlier reports in this set (Nos. 1 and 3) presented tabular results for the cœfficients of the terms independent of ϵ and linearly dependent on ϵ. The present report (in which the same value 1·405 for γ is used) deals with the cœfficients of ϵ^2 for cones of semi-angle between 5° and 25° [5(2·5)15, 20, 25] and gives in Part I tables of the additional second order effects on velocity, pressure and density. Part II contains information for deriving drag (which has a second order contribution in ϵ), normal force, ratio of shock wave to cone yaw, and the cœfficients determining the variations in entropy and in the shape of the shock wave which to this order ceases to be strictly circular in cross section. L. H.

The Origins of Modern Science. By Herbert Butterfield. (Bell & Sons.) Price 10*s*. 6*d*. net.

This work by the Professor of Modern History in the University of Cambridge is a welcome contribution to scientific literature.

A number of eminent scientists by their writings have demonstrated the cultural aspects of scientific thought and experiment. It is now gratifying to record the entry of a professional historian into this field in an endeavour to build a bridge between the humanities and the sciences.

In the reviewer's opinion the writer has achieved a conspicuous success. For this reason this is not the place to make minor criticisms. When the average student of science is so well versed in the history of his subject that he is capable of crossing swords with Professor Butterfield on his interpretation of some point in the views of, let us say, a sixteenth century physicist, then the bridge that the author is seeking to build will be truly fashioned. For this reason alone this volume is particularly to be commended to those who in Universities are concerned to see some broadening of the training of the specialized scientist.

An Advanced Treatise on Physical Chemistry.—Vol. I. *Fundamental Principles and Properties of Gases.* By J. R. Partington. [Pp. 943+40.] (Longmans.) Price 80*s*.

This book, as the author points out, is advanced in the sense of being comprehensive, and not in the sense of assuming much previous knowledge; for example, no previous knowledge of trigonometry or calculus is assumed, a concise account of these subjects being given in the mathematical introduction. This introduction is followed by sections on thermodynamics, kinetic theory of gases, statistical mechanics, quantum mechanics, thermometry, and a long final section on the properties of gases. Most sections start with a brief historical introduction and include a detailed account of the contributions of various workers in the field, with tables of experimental data and numerous references; the latter number about 18,000. Emphasis is laid throughout on the experimental side, though the theory is also thoroughly described. This book will be found a very valuable work of reference. A. F. D.

Applied Mathematics for Engineers and Scientists. By S. A. Schelkunoff. [Pp. 472.] (D. van Nostrand Company Inc., 1948.) Price 36*s*.

Present day science demands a mathematical discipline which often cannot be obtained from an undergraduate course. The author aims at providing this background for students of applied science. An idea of the ground covered can be obtained from the following list of contents: Complex numbers—Theory of approximation—Solution of equations—Power series—Differentiation—Integration—Vector analysis—Coordinate systems—Exponential functions—Differential equations of the first and higher orders—Partial differential equations—Conformal transformations—Contour integration—Linear analysis—The Gamma function—Exponential and Fresnel integrals—Bessel and Legendre functions—Formulation of equations. Thus the first half treats the general mathematical methods and the second half introduces the frequently occurring transcendental functions. The portion dealing with differential equations is particularly well written. The introductions to the use of Green's function and the Laplace transform method are two of the most valuable features of the book. The principles involved have been illustrated by concrete examples as far as possible. A chapter on tensors and matrices would perhaps have increased the general usefulness of the book. On the whole the book deserves to be widely used. **H. C. B.**

The Theory of Atomic Collisions. By N. F. Mott and H. S. W. Massey. Second Edition, Royal 8vo. [Pp. 388.] (Oxford: The University Press, 1949.) Price 35*s*. net.

In this work, which has been a classic since the publication of the first edition in 1933, the authors set themselves the task of discussing those collisions which do not involve creation or annihilation of particles, and can therefore be dealt with by wave-mechanics alone. In the second edition the form of the book remains unaltered, but the subject-matter is extended to include the phenomena of classical nuclear physics. There are valuable references to a large number of papers on both atomic and nuclear problems published since 1933, and the original treatment is expanded at varous points.

Broadly speaking, the book is divided into three parts. The first eight chapters are devoted to the general theory of collision processes, and include subsidiary mathematical and physical development where necessary. The main addition here is in the treatment of the dispersion formula, and of the collision complex or compound nucleus, although a more detailed analysis of the Born approximation has also been made.

Chapters IX. to XIII. deal with applications to atomic and nuclear collisions. More stress is now laid on nuclear problems and the chapter on the collision of electrons with molecules has been omitted.

In the last chapter relativistic two-body problems are considered. In orthodox quantum mechanics it has been usual to treat such problems by quantising the radiation field, thus bringing them outside the scope of this book, but the alternative method of the retarded potential has also been available, and the authors use this instead. Recently Feynman has criticised the use of field theory and developed the second method further, but it was obviously impossible to include a discussion of his work here. Some important work by Stueckelberg (1938), who foreshadowed Feynman's approach, has, however, apparently been overlooked by the authors. This chapter now includes a brief treatment of certain meson problems, together with useful references.

Altogether this new edition should prove invaluable to a wide class of physicists. **K. V. R.**

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

IX. *Atmospheric Absorption of Heat Radiation by Water Vapour.*

By T. G. Cowling*.

[Received September 23, 1949.]

Summary.

The absorption of heat radiation by water vapour in the atmosphere is considered. It is shown that existing experimental results are all consistent with Lorentz collision-broadening. A law is derived for the variation of absorption with amount of water vapour; this law differs somewhat from that proposed by Elsasser, indicating a slower extinction of radiation than the latter. The correction for pressure variations is found to be greater than Elsasser's normally; the temperature correction agrees with his.

§ 1. In an earlier report, the author has discussed the absorption of long-wave (heat) radiation by water-vapour in the atmosphere. The quantity of prime interest in that report was the fractional absorption A in relatively narrow wavelength ranges which are, none the less, sufficiently wide each to contain several absorption lines. The variation of A with the quantity w of absorbing vapour, with the pressure p, and with the temperature T, was given special attention. In the present paper, further experimental results regarding this variation are collected together, and are compared with the results of theoretical investigations.

The variation of A with w is found to follow a law somewhat different from that given by Elsasser (1942). It cannot be expressed by any simple analytical formula, and is best exhibited graphically. Plotting A directly against w is, however, inconvenient, since dA/dw rapidly falls off as w increases; it is preferable to plot A against $\log_{10} w$. The latter method of plotting, in addition to giving a more compact diagram, also has the advantage that an increase in the absorption coefficient in a definite ratio throughout the range leaves the form of the resulting curve unaltered, and simply displaces it bodily to left or right. This facilitates a comparison between the absorptions in wavelength ranges where the mean transparencies are very different, but where the types of absorption, as measured by the relative variations in the absorption coefficient throughout the ranges, may be expected to be similar.

The random effects of the distribution of strong absorption lines on the absorption curve are found to be smoothed out only if a frequency-range not much smaller than 25 cm.$^{-1}$ broad is used. At frequencies where the mean transparency does not vary greatly, the range can, however, be 100 cm.$^{-1}$ or more broad without any serious change in the absorption curve.

---

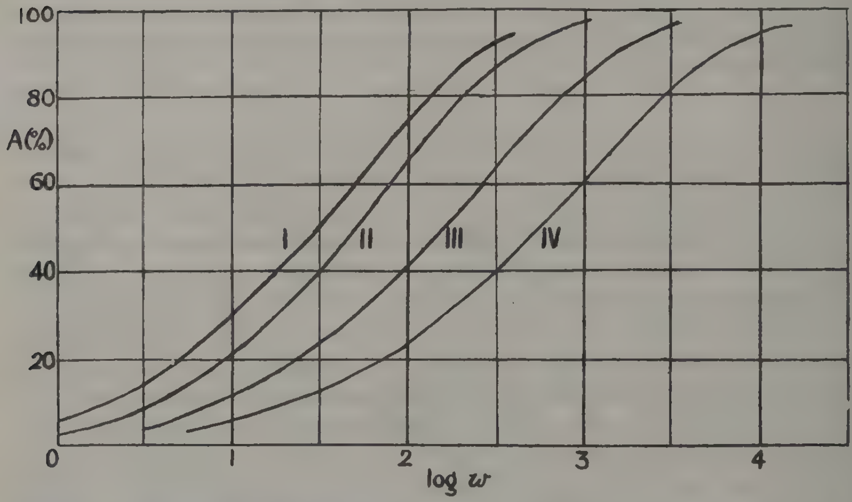
* Communicated by the Author.

§ 2. The following experimental studies have been made of the variation of A with w.

(i) The short-wave bands $\rho(0{\cdot}94\mu)$, $\Phi(1{\cdot}13\mu)$ and $\Psi(1{\cdot}38\mu)$ have been studied by various authors in order to determine spectroscopically the total amount of water vapour present in the atmosphere. In fig. 1, curve I., are shown Fowle's (1912) results for Ψ', giving the absorption in a range $0{\cdot}021\mu$ broad at $1{\cdot}47\mu$. His results for ranges of similar breadths in the bands ρ and Φ give curves very like that for Ψ'.

The results obtained by other authors for these bands will not be considered in detail. Herzing's (1937) results referred to wavelength ranges not small compared with the band width; Foster and Foskett (1945), on the other hand, appear to have used a very narrow range in studying the ρ-band (they do not state definitely its width, but speak of a range only a few Angstroms broad). Hand (1940) compared the absorption at the centre of the ρ-band with that at a neighbouring wavelength where the absorption, though much smaller, was not negligible.

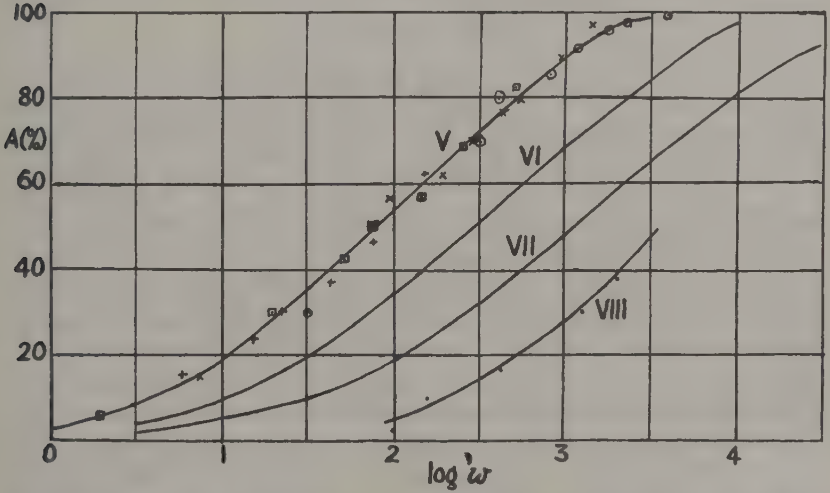
Fig. 1.



Graphs of absorption A against log w. Curve I., Fowle's observed results for Ψ'; curves II., III. and IV., calculated for the range 100–125 cm.$^{-1}$, with line-widths 0·4, 0·2 and 0·1 cm.$^{-1}$. The unit of w is, for curve I., 0·316 cm. precipitable water; for curve II., $1{\cdot}04\times10^{-5}$ cm.; for curve III., $5{\cdot}2\times10^{-6}$ cm.; and for curve IV., $2{\cdot}6\times10^{-6}$ cm.

(ii) The $6{\cdot}3\mu$ band has been studied by Fowle and Adel. Fowle (1917) made two bolograms (corresponding to $w=0{\cdot}008$ and $0{\cdot}082$ cm. of precipitable water) with a 60° rock-salt prism; his remaining bolograms were made with much smaller resolution, using a 15° prism. In the latter bolograms, all the bands at wavelengths below 5μ were completely smeared out; even beyond 5μ there was a considerable mixing of wavelengths which (because of the greater intensity of radiation in the shorter wavelengths) resulted in an apparent transfer of the absorption properties to higher wavelengths. We therefore use only the results of the 60° prism. These are combined with Adel's (1941) results for the transmission of sunlight through the atmosphere between $5{\cdot}5\mu$ and 14μ. Since absorption is almost complete between $5{\cdot}7\mu$ and 7μ, Adel's results give useful information only beyond 7μ. Near $7{\cdot}75\mu$, Adel's curves show the absorption of water vapour overlaid by that of N_2O, but it is possible to obtain a reasonable estimate of that part of the absorption which is due to water vapour.

Fig. 2.



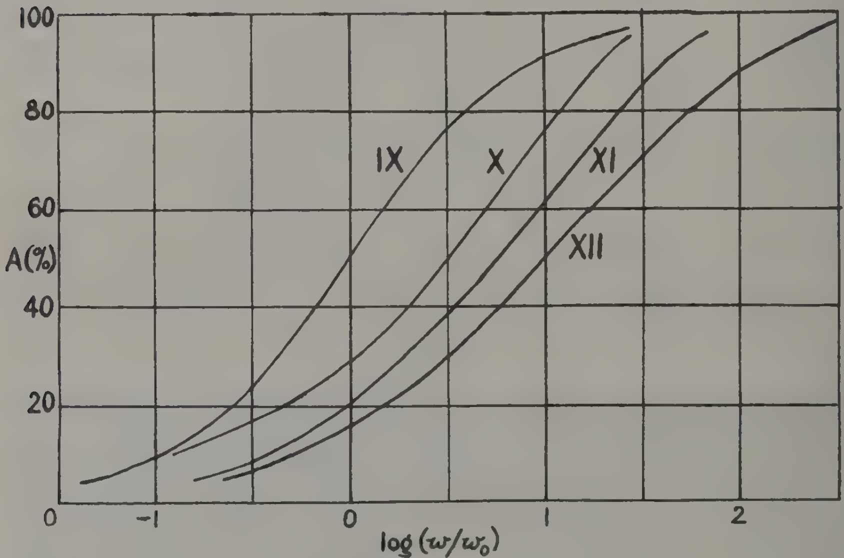
Graphs of absorption A against log w. Curves V., VI. and VII., calculated for the range 225–250 cm.$^{-1}$, with line-widths 0·4, 0·2 and 0·1 cm.$^{-1}$; curve VIII., derived from Strong's experimental results at 21μ (the actual experimental results are shown by dots). The unit of w is, for curve V., $1{\cdot}04\times10^{-5}$ cm.; for curve VI., $5{\cdot}2\times10^{-6}$ cm.; for curve VII., $2{\cdot}6\times10^{-6}$ cm.; and for curve VIII., 10^{-5} cm. Clustering round curve V. are shown Fowle and Adel's results for $7{\cdot}25\mu$ (⊙), $7{\cdot}5\mu$ (×), $7{\cdot}75\mu$ (⊡) and 8μ (+); the unit of w for these is, for $7{\cdot}25\mu$, $10^{-2\cdot6}$ cm.; for $7{\cdot}5\mu$, $10^{-1\cdot975}$ cm.; for $7{\cdot}75\mu$, $10^{-1\cdot4}$ cm.; and for 8μ, $10^{-0\cdot875}$ cm.

Plots of A against log w, using Fowle's and Adel's results, gave curves for 7·25, 7·5, 7·75 and 8μ which were almost identical in form. Using different units for w at the four wavelengths, the experimental points can be made to cluster round a single curve. They are shown in fig. 2 clustered round a curve (curve V.) whose theoretical derivation is described below (§ 4).

These experimental results, like those for the Ψ' band, are open to the objection that they refer, for large w, to transmission of sunlight through the atmosphere, whose pressure and temperature are non-uniform, This objection is, however, probably not serious, since most of the absorbing vapour is contained in the lowest km. or two of the atmosphere.

(iii) In the rotation band there are only isolated experimental results, save for a single series of data for 21μ obtained by Strong (1941) by residual ray methods. Strong has given results for shorter wavelengths also; but where data obtained by other methods are available, they are to be preferred. The residual ray method is hard to use, especially in regions of intense absorption where, owing to the length of the path traversed by radiation in the apparatus, it is difficult to estimate the intensity of radiation in the absence of absorption. Strong's data for 21μ are shown in fig. 2 (curve VIII.); the zero of absorption (not determined by him) is taken to correspond to a transmitted radiation equal to 1·95 of his units.

Fig. 3.



Graphs of absorption A against log (w/w_0). Curve IX. is Callendar's; curve X., Elsasser's; curve XI., Schnaidt's; and curve XII., that derived in the present paper. The values adopted for log w_0 for the four curves are 0, 0·5, 0·75 and 1.

§ 3. Next will be considered various theoretical and semi-empirical laws which have been proposed for the variation of A with w.

(i) Sutherland and Callendar (1943), following a discussion of CO_2 absorption by Callendar (1941), suggested that the infra-red absorption of most gases might be taken roughly to follow the empirical formula

$$A = w/(w+w_0), \quad \ldots \ldots \ldots \quad (1)$$

where w_0 is the value of w for which absorption is half complete. This is illustrated in fig. 3, curve IX. The formula gives only a moderate representation of Fowle's data for Ψ', and those of Fowle and Adel for $7{\cdot}25$–8μ; it indicates too rapid a variation of A with w below $A=50$ per cent, and too slow a variation above $A=80$ per cent. None the less, its simplicity makes it convenient for use in approximate work.

(ii) Elsasser (1938) derived a formula for the variation of A with w, assuming that the absorption lines are equal and evenly spaced, and that the absorption in each case is given by the Lorentz formula

$$k_\nu = k\gamma\pi^{-1}/\{(\nu-\nu_0)^2+\gamma^2\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

He found that, when absorption is complete at the line centres,

$$A = \operatorname{erf}(\delta^{-1}(\gamma k w \pi)^{\frac{1}{2}}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

where δ is the frequency-spacing between the lines. This formula is illustrated in fig. 3, curve X. When w is so small that absorption is incomplete at the line centres, the error-function law ceases to be applicable; it asserts that $A \propto w^{\frac{1}{2}}$ when w is small, whereas in fact $A \propto w$ then.

Comparison with the experimental results of Fowle and Adel not only confirms the inapplicability of the formula (3) for small w; it also shows that the error-function law gives too rapid an increase of A with w when $A>50$ per cent. The reason is to be found in the irregularity in strength and spacing of the absorption lines, which permits the existence of certain wide gaps between the stronger lines. Such gaps being relatively transparent, a much greater increase in w is needed to increase the total absorption from a low value to a high than would be required if the lines were regularly distributed.

(iii) Schnaidt (1939) determined the variation of A with w numerically, making assumptions similar to Elsasser's. A complete blackout of the line-centres was not assumed, so that Schnaidt's results for small absorptions effectively complete Elsasser's. Instead of considering the lines as overlapping, Schnaidt supposed each to be cut off at a distance from the line-centre equal to five times the line-width 2γ, the next line beginning there. This assumption is, of course, unjustified; but the resultant increase in transparency between the lines is much the same as would result from occasional wide gaps between the lines. In fact, Schnaidt's absorption curve (fig. 3, curve XI.) is very like Fowle's curve for Ψ' over the whole range of A.

§ 4. To test more exactly the effect of the irregularity of the lines, and of the incompleteness of absorption at the line-centres, I have determined numerically the law of variation of A with w for various ranges in the rotation band. The Lorentz formula (2) was assumed, and calculations were made with a line-width $2\gamma=0{\cdot}4$ cm.$^{-1}$ in the ranges 100–125, 225–250, 175–200, 200–225, 300–325 and 325–350 cm.$^{-1}$; also with line-widths of 0·8, 0·2 and 0·1 cm.$^{-1}$ for the first two ranges. The strengths and positions of the absorption lines were taken from the work of Randall, Dennison, Ginsberg and Weber (1937). The calculations were mainly done in 1943; had they been done more recently, smaller line-widths might have been used, and the strengths of absorption lines might have been varied to accord with the work of Cross, Hainer and King (1944). However, the similarity between the results from the different ranges, in which the distribution of line-strengths was very different, indicated that a moderate variation in line-strengths cannot seriously affect the absorption curves.

In the calculations with $2\gamma=0\cdot4$ cm.$^{-1}$, the total absorption coefficient from all neighbouring lines was first found at intervals of $0\cdot2$ cm.$^{-1}$ throughout the range under consideration. The next step was to find the fractions of the range for which the logarithm (to base 10) of the absorption coefficient lay in each of successive ranges of magnitude 0·05. The contribution to the total absorption from these fractions were then determined for values of $\log_{10} w$ at intervals of 0·25, and were summed. Similar methods, with a different frequency-spacing, were used for the other line-widths.

The results for $2\gamma=0\cdot4$, $0\cdot2$ and $0\cdot1$ cm.$^{-1}$ are shown in fig. 1 (curves II., III., IV.) for the range 100–125 cm.$^{-1}$, and in fig. 2 (curves V., VI., VII.) for 225–250 cm.$^{-1}$. The second set of curves shows a slower increase of A with w than the first; this is due to the wide gap containing no absorption lines between 223·4 and 244·5 cm.$^{-1}$, which gives appreciable absorption only when w is large. Gaps like this, though usually less wide, occur frequently as the wing of the band is approached. The absorption curves consequently tend to show a decrease in steepness as the wave-number increases, though the curve for 225–250 cm.$^{-1}$ actually represents an extreme form of band-wing absorption; the curves for 175–200 and 200–225 cm.$^{-1}$ have practically the same shape as that for 100–125 cm.$^{-1}$, while those for 300–325 and 325–350 cm.$^{-1}$ have shapes intermediate between this and that for 225–250 cm.$^{-1}$. A reduction in the line-width also leads to a decreased slope of the absorption curve.

Curves I. and II. of fig. 1 are very similar in shape. This is satisfactory, for it provides a direct experimental check on the theoretical curves. Cornell (1937) found mean line-widths 0·57 and 0·4 cm.$^{-1}$ for the ρ and Φ bands respectively; thus the line-width in the Ψ' band is also probably about 0·4 cm.$^{-1}$. Hence if the distribution of strong and weak lines has the same random characteristics in the Ψ' and rotation bands, the absorption curves for the two should be much the same. The fact that this is indeed so is a striking confirmation of the validity of the Lorentz formula (2), used in the calculations.

The experimental results for $7\cdot25$–8μ are shown in fig. 2 clustering closely round curve V. The agreement with this curve supplies a further experimental check on theory. The experimental results, like the theoretical curve, represent band-wing absorption ($7\cdot25\,\mu$ is about 215 cm. ^{1}

from the centre of the $6{\cdot}3\mu$ band). The closeness of agreement between the two suggests that the line-width in the $6{\cdot}3\mu$ band is not very different from 0·4 cm.$^{-1}$, but little weight can be attached to this suggestion; one would obtain a similar absorption curve, assuming a less extreme band-wing absorption but a smaller line-width.

The line-width at the wing of the rotation band was found by Adel (1947) to be 0·22–0·24 cm.$^{-1}$. There is no essential reason why the line-width should remain the same right down to very low wave-numbers, but observations at centimetre wave-lengths actually show a line-width which is similar in order of magnitude, though a little smaller (Becker and Autler 1946). Taking a uniform line-width $2\gamma=0{\cdot}24$ cm.$^{-1}$, the numerical calculations just described can be used to estimate the actual absorption in various parts of the rotation band, as well as the run of the absorption curve. A comparison of the calculated absorptions with such scanty experimental data as exist provides a further check on theory.

In the range 100–125 cm.$^{-1}$, Wright and Randall (1933) give transmission curves for 80 cm. of damp air, from which I estimate an absorption of roughly 55 per cent. The humidity is not quoted, but the quantity of absorbing water may be estimated as $(7\pm2)\times10^{-5}$ cm.; for this, the calculated absorption is 50 ± 6 per cent. In the ranges 300–325 and 325–350 cm.$^{-1}$, Hettner (1918) found $2{\cdot}5\times10^{-3}$ cm. of precipitable water to give 52 per cent and 49 per cent absorption respectively. This was at 354° K.; I estimate that $2{\cdot}5\times10^{-3}$ cm. at this temperature should produce the same absorption, in the two ranges, as about 3×10^{-3} and $3{\cdot}4\times10^{-3}$ cm. at room temperature. The theoretical absorptions of these amounts of water are 45 per cent and 43 per cent. In the ranges 200–225 and 175–200 cm.$^{-1}$, Kühne (1933) found 77 per cent and 69 per cent absorption to be produced by about $3{\cdot}3\times10^{-3}$ cm. of water. The corresponding theoretical absorptions are 84 per cent and 78 per cent.

In view of the uncertainties, these figures show fair agreement between theory and experiment. The theory is admittedly approximate: the line-strengths are somewhat uncertain, particularly in the band-wing, where centrifugal stretching of the molecules may be important. Nevertheless, a comparison between curve VIII. of fig. 2 and the other curves of this figure indicates no reason to believe that the shape of the absorption curve at the extreme wing of the band is different from that at greater wavelengths.

It appears, then, that a single absorption curve can be used for regions near a band-centre, and a second for band-wings; because of possible differences in line-width, different curves should be used for the $6{\cdot}3\mu$ and rotation bands. However, in atmospheric work complication is avoided, and remarkably little error is involved, if a single absorption curve is used at all wavelengths. Such a curve is shown in fig. 3 (curve XII.); the material from which it is constructed is given in Table 1. The table is based on the curves of figs. 1 and 2, adopting a line-width rather above $2\gamma=0{\cdot}25$ cm.$^{-1}$, and a type of absorption approximating more closely to the band-centre absorption of fig. 1 than to the rather extreme band-wing absorption of fig. 2. Tables are also given for line-widths one-half and one-quarter as large, corresponding to pressures of one-half and one-quarter of an atmosphere.

TABLE I.

Percentage absorption for given $\log_{10}(w/w_0)$, at pressures of 1, ½ and ¼ atmosphere.

$\log_{10}(w/w_0)$	Absorption 1 atm.	½ atm.	¼ atm.	$\log_{10}(w/w_0)$	Absorption 1 atm.	½ atm.	¼ atm.
−0·2	42	42½	43½	0·2	58½	58	57½
−0·4	34	35½	36½	0·4	67	66	65
−0·6	27	29	30½	0·6	75	73½	72½
−0·8	20½	22	23	0·8	82½	81	79½
−1·0	15½	17½	19	1·0	88	86½	85
−1·2	11½	13½	14½	1·2	93½	92	91
−1·6	5½	7½	8½	1·4	97	96	95
−2·0	2½	4	5½	1·6	99	98½	98
−2·4	1	2	3	1·8	100	100	99½

TABLE II.

Values of w_0, for which absorption is half complete, at room temperature and at 220° K. (w_0 in cm. precipitable water).

Range	$\log_{10} w_0$ Room temperature	220°	Range	$\log_{10} w_0$ Room temperature	220°
5·0 –5·2μ	−1·3	−1·15	8–13μ	0·6	0·85?
5·2 –5·4	−1·6	−1·55	13–15	0·0?	0·3 ?
5·4 –5·6	−1·9	−1·9	15–17	−0·5?	0·0 ?
5·6 –5·8	−2·35	−2·35	17–19	−1·0	−0·35
5·8 –6·05	−2·6	−2·65	19–21	−1·4	−0·9
6·05–6·2	−2·3	−2·35	21–23	−1·75	−1·3
6·2 –6·35	−1·95	−2·0	23–25	−2·0	−1·6
6·35–6·45	−2·5	−2·55	25–27	−2·2	−1·85
6·45–6·7	−2·65	−2·7	27–30	−2·4	−2·15
6·7 –6·9	−2·2	−2·2	30–34	−2·6	−2·45
6·9 –7·2	−1·85	−1·85	34–40	−2·8	−2·7
7·2 –7·4	−1·45	−1·4	40–50	−3·0	−3·0
7·4 –7·6	−1·05	−0·95	50–∞	−3·1	−3·15
7·6 –7·8	−0·6	−0·4			
7·8 –8·0	−0·15	0·1			

Table I. gives A for various values of $\log_{10}(w/w_0)$, where w_0 is, as before, the value of w for which A is 50 per cent. It needs to be supplemented by a further table giving w_0 for different wavelength ranges; this is given, for a pressure of one atmosphere, in Table II., columns (2) and (5).

The experimental work on which this table is based is, in general, that used in constructing Table II. of the earlier report (Cowling 1943); beyond 34 μ, however, the numerical calculations described above have been used, and the residual ray observations of Strong and Woo (1932) have been disregarded. In the range 8–13μ, though values of w_0 are given, the absorption may approximate more closely to continuous absorption than to the kind of band absorption to which Table I. applies. The curves of figs. 1 and 2 enable one to estimate that the values of $\log_{10} w_0$ at pressures of $\frac{1}{2}$ and $\frac{1}{4}$ atmos. respectively are 0·25 and 0·51 greater than those at atmospheric pressure.

§ 5. *Pressure effects.* In the above it has been assumed, in accordance with the Lorentz theory of collision-broadening, that a reduction in pressure leads to a reduction in line-width in the same ratio. The Lorentz theory is usually accepted as valid for the infra-red absorption of gases (see, *e. g.*, Coggeshall and Saier 1947). Nevertheless, some workers have doubted its applicability to water vapour absorption, for the following reasons.

The absorption of water vapour, when between about 10 per cent and 50 per cent complete, approximately obeys the square-root law $A \propto w^{\frac{1}{2}}$ (see, *e. g.*, Strong 1941). This law can be given a theoretical justification, if the Lorentz formula (2) is assumed to be obeyed, and the amount of water present is supposed to be enough to black out the centres of absorption lines, but insufficient to produce much absorption in their overlapping wings. Making these assumptions, it is found that

$$A=2\delta^{-1}w^{\frac{1}{2}}\Sigma(\gamma k)^{\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where δ is the frequency-range considered, and the summation is over all lines whose centres are in the range. Since γ is, on the collision-broadening theory, proportional to the pressure p, (4) implies that $A \propto p^{\frac{1}{2}}$; that is, a reduction in p produces the same decrease in A as a reduction in w in the same ratio. However, experiments show that, even for values of w such that $A \propto w^{\frac{1}{2}}$, A varies much more slowly than $p^{\frac{1}{2}}$. Falckenberg (1938) found that the effect on the total absorption, integrated over all wavelengths, of a k-fold decrease in p, is roughly equivalent to that of reducing w in the ratio $k^{\frac{1}{2}}$; Schnaidt (1939) obtained a similar result from v. Bahr's (1909) experimental results for absorption near 2·7μ. Though their results referred to a very limited range of variation of w, this law of pressure variation of the absorption has often been taken as generally valid, and has been interpreted as implying that the line-width is not proportional to the pressure itself, but to its square root.

But it is not actually necessary to abandon the Lorentz theory. The numerical work described in the last section enables one to see whence the apparent contradiction has arisen. It shows that there are no very secure theoretical grounds for using the formula (4). No range of values of w exists such that the absorption at line-centres is complete, and that in the region of overlapping wings weak; the wings of strong lines begin to encroach on neighbouring lines before the centres of weak lines are blacked out. The calculations for the ranges 100–125 and 225–250 cm.$^{-1}$, with line-width $2\gamma=0{\cdot}2$ cm.$^{-1}$, give absorptions such that $A\propto w^{\frac{1}{2}}$, roughly, when A is about 25 per cent ; but the actual absorption is then never greater than $\frac{7}{8}$ that given by (4), even when in using this equation some allowance is made for overlapping by lumping together all lines whose centres are less than 0·5 cm.$^{-1}$ apart.

The numerical work equally shows that, on the Lorentz theory, the pressure variation of A is never as fast as proportional to $p^{\frac{1}{2}}$. A reduction of 2γ from 0·2 to 0·1 cm.$^{-1}$ (equivalent to halving the pressure) leads to a decrease in A, in the same two frequency-ranges, which is something like in the ratio 0·8 when A lies between 10 and 50 per cent ; the decrease is less when A lies outside these limits. The effect on A of a k-fold decrease in p is, moreover, not equivalent to that of a $\sqrt{k}$-fold decrease in w ; when A is less than 5 per cent a reduction in p produces little change in A, whereas if A exceeds 60 per cent a reduction in p has nearly the same effect as a proportionate reduction in w.

§ 6. *Integral absorptions.* These remarks refer to the absorption in relatively narrow wavelength ranges. Tables I. and II. can also be used to determine the integral absorption of radiation from a black body, given by

$$\overline{A}=\Sigma f A, \quad \ldots\ldots\ldots\ldots \quad (5)$$

where f is the fraction of the black-body radiation in the range in which the absorption is A. If the absorbing vapour is at the same temperature as the black body, $\overline{A}$ is, of course, the emissivity of the vapour, expressed as a fraction of the total black-body radiation.

I have determined the values of $\overline{A}$ numerically for a black body at 286° K. at pressures of one, a half, and a quarter atmosphere. The values are somewhat uncertain for large values of w, for which absorption in the relatively transparent range 8–13μ is important. Absorption in this range does not appear to vary with w in the way illustrated by Table I. Its run at atmospheric pressure can be estimated from Adel's (1941) diagrams ; values at other pressures are uncertain since, as Elsasser suggests, the continuous part of the absorption here may be due to the wings of far distant absorption lines, with an absorption coefficient inversely proportional to the pressure. Such uncertainties affect the values of $\overline{A}$ appreciably when w exceeds a few mm. of precipitable water, and so the calculations for the smaller pressures have been restricted to small quantities of water. For such values of w, the error in $\overline{A}$ is not likely to exceed 1 per cent or so of the black-body radiation, and even when $\overline{A}$ is greater than a few mm. of water the error in the values at atmospheric pressure is not likely to be much larger.

The calculations indicate a pressure dependence of A which is appreciable even when $\overline{A}$ is as small as 2 per cent (this corresponds to A being about

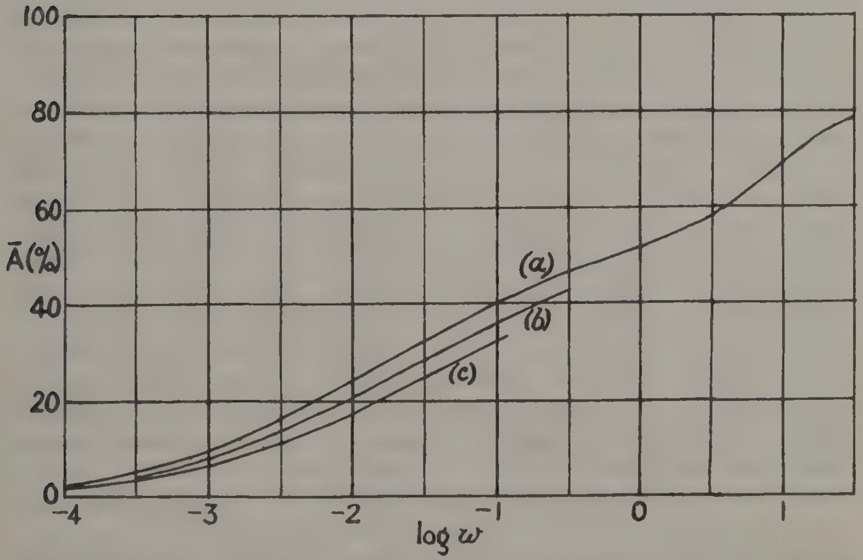
15 per cent in the more opaque wavelengths.) A four-fold reduction in pressure is equivalent to replacing w by $\frac{2}{3}w$ when $\bar{A}=2$ per cent, by $\frac{1}{2}w$ when $\bar{A}=9$ per cent, and by $\frac{1}{3}w$ when $\bar{A}=40$ per cent. The calculated values for $w=1{\cdot}2\times10^{-3}$ and $2{\cdot}4\times10^{3}$ cm. are compared with Falckenberg's (1938) experimental values in Table III. The two sets of values indicate

TABLE III.

Calculated integral absorptions, compared with Falckenberg's observed values.

	$w=1{\cdot}2\times10^{-3}$ cm.		$w=2{\cdot}4\times10^{-3}$ cm.	
	$\bar{A}$ (obs.) (per cent)	$\bar{A}$ (calc.) (per cent)	$\bar{A}$ (obs.) (per cent)	$\bar{A}$ (calc.) (per cent)
1 atm.	7·8	10·8	12·4	14·6
½ atm.	6·4	8·7	—	—
¼ atm.	4·5	6·9	8·4	9·6

Fig. 4.



Graphs of integral absorption $\bar{A}$ against log w (w measured in cm.) at pressures of 1 (curve (*a*)), ½ (curve (*b*)), and ¼ (curve (*c*)) atmosphere.

pressure changes in the absorption in roughly the same ratio. The actual values of the experimental absorption do not agree at all closely with theory, but this is not surprising. The experimental results refer to absorption of radiation from a source at 100° C.; as Schnaidt (1938) pointed out, they need to be corrected by increasing them by about 30 per cent; and if they are corrected thus, the values at 1 atmos. are definitely higher than those found by other workers.

The calculated values of $\bar{A}$ at the different pressures are shown in fig. 4. Absorption in the ranges 0–5μ and 13–17μ is omitted, since in the atmosphere water-vapour absorption in these ranges is dominated by that of CO_2; this facilitates a comparison with experiments made with atmospheric air and corrected by subtracting the estimated CO_2 absorption. For small values of w (less than $10^{-2\cdot5}$ cm.) the calculated $\bar{A}$ at atmospheric pressure is about one-sixth less than the emissivity found by Falckenberg (1939), and about the same amount greater than that found by Elsasser (1941); it agrees fairly closely with that found by Brooks (1941). For large values of w (greater than 1 mm.) the calculated $\bar{A}$ is less than Elsasser's by two or three per cent of the black-body radiation, and a similar amount less than of Brooks: it exceeds by a rather smaller amount that of Robinson (1947), derived from what he calls observations of minimum radiation, but agrees with the mean of Robinson's observations, including those which he regards as subject to extraneous radiation. In view of the uncertainties the agreement of the calculated absorptions with experiment is satisfactory.

§ 7. *Atmospheric pressures effects.* By reason of the non-uniformity of pressure the problem of pressure corrections to atmospheric absorption is more subtle than the corresponding problem when the whole absorbing column is at the same pressure. Consider, for example, absorption of radiation leaving the earth's surface in a given direction and traversing the whole atmosphere. The pressure correction to the part of the absorption which takes place in the upper atmosphere has to be applied to an *increment* to the absorption. At different pressures the separate wavelengths of a range contribute slightly different amounts to a total absorption A of assigned amount in that range. Ignoring this difference, which is probably unimportant save for very small values of A, the pressure correction to an increment of A is equivalent to reducing the amount of water vapour in the ratio $(dA/dw)_p : (dA/dw)_a$, where $(dA/dw)_a$ is the value of dA/dw at atmospheric pressure and $(dA/dw)_p$ is that at the pressure concerned. The value of the ratio depends on A, the absorption in layers below that considered. A similar correction has to be applied in considering radiation reaching the ground surface from the upper atmosphere.

Values of the ratio, deduced from the data of figs. 1 and 2, are given in Table IV. It will be noted that the pressure correction is greater than reducing the quantity of water vapour in the same ratio as $p^{\frac{1}{2}}$; save for very small absorptions the reduction is more nearly in the same ratio as p. Table IV. also contains values of $(d\bar{A}/dw)_p : (d\bar{A}/dw)_a$, giving the similar reduction in the effective water amount in considering the integral absorption; the values of this ratio vary less with the total absorption than do those of the first ratio.

§ 8. *Temperature effects.* A variation in the temperature T affects the absorption in two ways. At given pressure, the line-width varies like $T^{-7/10}$. Temperature changes also affect the populations of the different energy-levels; the high-energy levels, which contribute most to band-wing absorption, are but scantily occupied at low temperatures. A decrease in temperature therefore leads to an increase in absorption near a band centre (for the rotation band at great wavelengths) but makes the absorption cut-off at the wings of the band sharper.

The only genuinely reliable theoretical method of estimating the absorption at low temperatures would be by carrying out calculations like those described in § 4, for each of a series of wavelength ranges. A less complicated approximate method has, however, been used. The intensities of lines in the rotation band between 75 and 550 cm.$^{-1}$ were first calculated for a variety of temperatures. As the temperature falls off, weak lines are found to be reduced in strength more than the strong ones; thus the gaps between strong lines tend to be more accentuated and the absorption tends to possess more of the characteristics of band-wing absorption. The effect is not large, however, and is offset by the increased line-width; hence it is reasonable to take A to vary with w according to Table I., as at ordinary temperatures. Then, to determine the absorption at any wavelength, it is necessary only to know w_0.

Table IV.

Values of $(dA/dw)_p/(dA/dw)_a$ and $(d\bar{A}/dw)_p/(d\bar{A}/dw)_a$, corresponding to $p=\frac{1}{2}$ and $\frac{1}{4}$ atmosphere.

	$\left(\frac{dA}{dw}\right)_p : \left(\frac{dA}{dw}\right)_a$			$\left(\frac{d\bar{A}}{dw}\right)_p : \left(\frac{d\bar{A}}{dw}\right)_a$	
A (per cent)	$\frac{1}{2}$ atm.	$\frac{1}{4}$ atm.	$\bar{A}$ (per cent)	$\frac{1}{2}$ atm.	$\frac{1}{4}$ atm.
5	0·82	0·6	5	0·7	0·44
10	0·72	0·46	10	0·65	0·39
15	0·67	0·4	15	0·62	0·36
20	0·64	0·36	20	0·60	0·34
30	0·60	0·33	25	0·58	0·33
40	0·57	0·3	30	0·56	0·31
50	0·54	0·28	35	0·55	0·30
60	0·52	0·265	40	0·55	—
80	0·5	0·25			

To determine w_0 the formula (4) is used. This formula indicates 19 per cent absorption when the true absorption is about 15½ per cent. Hence w_0 is about ten times the value of w which (4) indicates to give 19 per cent absorption.

In using (4) one is, of course, only following Elsasser (1938), who calculated the coefficient of $w^{\frac{1}{2}}$ in this formula for ranges 25 cm.$^{-1}$ broad between 75 and 550 cm.$^{-1}$, at temperatures between 200 and 320° K. As noted in my earlier report, Elsasser's calculations were subject to systematic errors; my own results agree with his (as regards order of magnitude) between 75 and 375 cm.$^{-1}$, but give values of w_0 only one-hundredth as large as his in the range 475–550 cm.$^{-1}$. Nevertheless, my results differ little from his in the ratios which they give for the values of w_0 at different temperatures for any given wavelength range, and it is therefore unnecessary to give my results in detail. They indicate that, for wavelengths greater than 50 μ, the absorption at 220° K. is slightly greater than that at room temperature; in the range 40–50 μ the absorptions at the two temperatures are roughly the same; while for shorter wavelengths the absorption at 200° K. becomes progressively the smaller of the two, until near 20 μ the value of w_0 at 220° K. is about five times that at room temperature. The estimate at 20 μ cannot be regarded as very reliable; the estimated value of w_0 at room temperature here is at least twice that indicated by experiment. However, the calculated ratios of w_0 at room temperature and at 200° K. have been used to construct a table for w_0 at 220° K.; this is given in Table II., third and sixth columns. The temperature variation at a frequency ν in the 6·3 μ band is taken to be roughly that at a frequency $|\nu-\nu_0|$ in the rotation band, where ν_0 is the frequency of the 6·3 μ band centre.

TABLE V.

Values of $\overline{\mathrm{A}}$ (a) at 220° K. and $\frac{1}{4}$ atm., compared with (b) values uncorrected for the change of absorption with temperature, and (c) values at room temperature and pressure $\frac{1}{4}$ atm.

w (cm.)	10^{-4}	$10^{-3 \cdot 5}$	10^{-3}	$10^{-2 \cdot 5}$	10^{-2}	$10^{-1 \cdot 5}$	10^{-1}
percentage $\overline{\mathrm{A}}$ (a)	2·3	4·5	8·2	13·6	20·65	28·9	37·4
percentage $\overline{\mathrm{A}}$ (b)	2·6	5·3	9·7	16·0	24·2	33·7	43·0
percentage $\overline{\mathrm{A}}$ (c)	1·7	3·4	6·4	11·1	17·2	24·8	32·5

The temperature variation in w_0 is large only in the relatively transparent band wings. In the upper atmosphere, where the total amount of absorbing vapour is small, its effect on the integral absorption $\overline{\mathrm{A}}$ is masked by the increasing importance, at low temperatures, of long wavelengths, where water vapour absorbs strongly. Table V. gives the calculated integral absorption at 220° K. of black-body radiation at this temperature, both when account is taken of the temperature variation of w_0, and when the values of w_0 appropriate to room temperature are used uncorrected. The calculations refer in each case to a pressure of $\frac{1}{4}$ atmos.; the corresponding integral absorptions at room temperature are also given for comparison. The increase in $\overline{\mathrm{A}}$ due to the increased importance of long wavelengths at low temperatures is about twice the decrease due to the increased transparency of the band-wings.

§ 9. *Conclusion.* In discussing the absorption of long-wave radiation in the atmosphere, drastic simplifying approximations are often necessary. It is hoped that the data given in this paper will enable the value of such approximations to be assessed. For example, the pressure correction which replaces a quantity w of water vapour at pressure p by a quantity $w\sqrt{(p/p_a)}$ at atmospheric pressure p_a is in most cases an undercorrection : the true correction is often nearer to replacing w by wp/p_a.

However, it must be emphasized that the approximations justified in any particular problem must depend on that problem. As noted in the last section, the emissivity of a small quantity of water vapour, measured as a fraction of the black-body radiation, increases with decreasing temperature. But the radiation reaching the ground from great heights does not mirror this increased emissivity. The radiation emitted by water at great heights is chiefly emitted in very long wavelengths, which are subject to great absorption in the lower levels. On the other hand, in considering the radiative equilibrium of the stratosphere the increased emissivity at low temperatures should be important.

REFERENCES.

ADEL, A., 1941, *Astrophys. J.*, **94,** 451 ; 1947, *Phys. Rev.*, **71,** 806.

v. BAHR, E,. 1909, *Ann. der Phys.* [4], **55,** 476.

BECKER, G. E., and AUTLER, S. H., 1946, *Phys. Rev.*, **70,** 300.

BROOKS, F. A., 1941, *Mass. Ints. Tech. Papers in Phys. Oceanography and Met.*, **8,** no. 2.

CALLANDER, G. S., 1941, *Qu. J. Roy. Met. Soc.*, **67,** 263.

COGGESHALL, N. D., and SAIER, E. L., 1947, *J. Chem. Phys.*, **15,** 868.

CORNELL, S. D., 1937, *Phys. Rev.*, **51,** 739.

COWLING, T. G., 1943, *Rep. Prog. Phys.*, **9,** 29.

CROSS, P. C., HAINER, R. M., and KING, G. W., 1944, *J. Chem. Phys.*, **12,** 210.

ELSASSER, W. M., 1938, *Astrophys. J.*, **87,** 496 ; 1941, *Month. Weather Rev.*, **69,** 1. ; 1942, *Harvard Meteorological Studies*, no. 6

FALCKENBERG, G., 1938, *Met. Z.*, **55,** 174 ; 1939, *Ibid.*, **56,** 72.

FOSTER, N. B., and FOSKETT, L. W., 1945, *J. O. S. A.*, **35,** 601.

FOWLE, F. E., 1912, *Astrophys. J.*, **35,** 149 ; 1917, *Smithson. Misc. Coll.*, **68,** no. 8.

HAND, I. F., 1940, *Month. Weather Rev.*, **68,** 95.

HERZING, F., 1937, *Gerl. Beit.*, **49,** 71.

HETTNER, G., 1918, *Ann. der Phys.*, **55,** 476.

KUHNE, J., 1933, *Z. f. Phys.*, **84,** 722.

RANDALL, H. M., DENNISON, D. M., GINSBERG, N , and WEBER, L. R., 1937, *Phys. Rev.*, **52,** 160.

ROBINSON, G. D., 1947, *Qu. J. Roy. Met. Soc.*, **73,** 127.

SCHNAIDT, F., 1939, *Gerl. Beit.*, **54,** 203.

STRONG, J., 1941, *J. Franklin Inst.*, **232,** 1.

STRONG, J., and WOO, S. C., 1932, *Phys. Rev.*, **42,** 267.

SUTHERLAND, G. B. B. M., and CALLANDER, G. S., 1943, *Rep. Prog. Phys.*, **9,** 29.

WRIGHT, N., and RANDALL, H. M., 1933, *Phys. Rev.*, **44,** 391.

X. *Orthogonal Polynomials in the Least Squares Fitting of Observations.*

By P. G. Guest,
Lecturer in Physics, University of Sydney *.

[Received May 25, 1949.]

Summary.

The theory of least squares curve-fitting of non-equidistant data is developed in terms of orthogonal polynomials and power moments. A modification of the Doolittle scheme is proposed which enables probable error calculations to be carried out very simply. The modified scheme is illustrated by a specific example.

§ 1. Introduction.

In recent years the use of orthogonal polynomials in the fitting of curves to equally-spaced observations has been thoroughly investigated by a number of authors (Birge and Shea 1927, Aitken 1932, Davis 1935, Kerawala 1941, Fisher 1946, Birge 1947). A comprehensive discussion of these investigations has been given by Birge (1947).

When the observations are not made at equally-spaced intervals, the special methods of these authors are no longer applicable. In these general cases it is necessary to solve the normal equations algebraically. The usual method employed is that of Doolittle (1878). In the Doolittle method the calculation of the power-series coefficients is quite straightforward, but the probable error calculations (Brunt 1917) are somewhat involved.

Probable error calculations are greatly simplified by the use of orthogonal polynomials. Fisher (1946) states that the method of orthogonal polynomials becomes artificial in the general case where the observations are not equally spaced. This is true in the sense that the calculation of the orthogonal polynomials $T_j(x)$ at each point of observation for use in the orthogonal moments $\Sigma y T_j(x)$ would require an exorbitant amount of time. In fact, as Aitken (1933–34) points out, the only practicable method is to base the calculations on the power moments $\Sigma y x^j$.

Although the direct calculation of orthogonal polynomials and orthogonal moments is out of the question, it is highly desirable that the fitted curve should be expressed in orthogonal as well as in power-series form. The advantages of orthogonal representation are :—

(i) All curves of degree less than p are obtained in the process of obtaining the curve of degree p.

(ii) The probable error calculations for the power-series coefficients and for the fitted values are greatly simplified.

---

* Communicated by the Author.

The purpose of the present paper is to develop the general theory of curve-fitting in terms of orthogonal polynomials and power moments. By this means the advantages of the orthogonal method of solution are obtained without the prohibitive labour of calculating directly the orthogonal moments.

It is found that the arithmetical procedure is very closely related to that of the Doolittle scheme. The various terms in the Doolittle scheme are shown to be identical with terms appearing in the expansion of the orthogonal polynomials. The values $S_{jj}(=\Sigma T_j^2(x))$ and β_{jk} (identical with Birge's R_{jk}) which form the basis of the probable error calculations may in fact be calculated from terms which appear in the Doolittle scheme.

A variation of the Doolittle scheme is proposed which makes use of the advantage arising from orthogonal representation. The scheme is extended to include the calculation of probable errors both of the power-series coefficients and the fitted values.

Aitken (1932–34) in a series of three papers has given a comprehensive treatment of the problem of curve-fitting. In the second of these papers he develops a general scheme applicable to the case of non-equidistant data. It is doubtful whether Aitken's scheme would be of very great practical use except where the values Σx^j are fairly small. The scheme unlike that of Doolittle, calculates the coefficients β_{jk}, and so can be extended to include probable error calculations. It is considered briefly in §2.

Tchebycheff originally obtained his polynomials in the form of continued fractions. Isserlis (1927) has discussed the problem of curve-fitting by means of Tchebycheff polynomials expressed in this form. Though interesting from a mathematical viewpoint, his scheme is too involved to yield a satisfactory practical procedure.

§ 2. General Theory.

Let $y(x_i)$, $i=1, 2, \ldots n$, represent the n observations which are to be fitted to a polynomial of degree p in x, the errors in the observed values of the variable x being supposed negligible. The polynomial $u_p(x)$ to fit these observations is to be chosen so that the sum of the squares of the residuals $y(x_i)-u(x_i)$ should be a minimum.

(a) *Calculation of Tchebycheff Polynomials.*

The polynomial $u_p(x)$ will be expanded in terms of polynomials $T_j(x)$ of degree $j=0$ to p, which are orthogonal over the particular set of values x_i at which the observations were made. The polynomials $T_j(x)$ are the orthogonal polynomials of Tchebycheff, satisfying the equations

$$\sum_{i=1}^{n} T_j(x_i)T_k(x_i)=0, \quad j \gtrless k. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The equations (1) determine the polynomials $T_j(x)$ except for a single arbitrary constant in each polynomial. This constant is accounted for by imposing the additional condition that the coefficient of the leading term x^j in the power-series expansion of $T_j(x)$ should be unity.

In order to derive an explicit expression for the polynomial $T_j(x)$, it is written as a linear function of the polynomials of lower degree.

$$T_j(x)=x^j+\alpha_{j,j-1}T_{j-1}(x)+\alpha_{j,j-2}T_{j-2}(x)+\ldots+\alpha_{j0}. \tag{2}$$

To find the coefficient α_{jk}, (2) is multiplied by $T_k(x)$ and summed over the values x_i. Thus

$$\Sigma T_j(x)T_k(x)=\Sigma x^j T_k(x)+\alpha_{jk}\Sigma T_k^2(x)=0$$

and

$$\alpha_{jk}=-\Sigma x^j T_k(x)/\Sigma T_k^2(x).$$

This can be written as

$$\alpha_{jk}=-S_{jk}/S_{kk}, \tag{3}$$

where

$$\left.\begin{aligned} S_{jk}&=\Sigma x^j T_k(x),\\ S_{kk}&=\Sigma x^k T_k(x)\equiv\Sigma T_k^2(x).\end{aligned}\right\} \tag{4}$$

Now

$$\begin{aligned} S_{jk}&=\Sigma x^j T_k(x)\\ &=\Sigma x^j(x^k+\alpha_{k,k-1}T_{k-1}(x)+\ldots+\alpha_{k0})\\ &=S_{j+k,0}+\alpha_{k,k-1}S_{j,k-1}+\ldots+\alpha_{k0}S_{j0}. \end{aligned} \tag{5}$$

The values of S_{jk} and α_{jk} can then be calculated in succession from the sums of the powers $S_{j+k,0}$ and the previously calculated values S_{jm} and α_{km} $(m<k)$.

To express $T_j(x)$ as a power series

$$T_j(x)=x^j+\beta_{j,j-1}x^{j-1}+\ldots+\beta_{j0}, \tag{6}$$

the coefficients β_{jk} are determined by expanding each term on the right-hand side of (2) in powers of x and collecting terms in x^k. Thus

$$\beta_{jk}=\alpha_{jk}+\alpha_{j,k+1}\beta_{k+1,k}+\alpha_{j,k+2}\beta_{k+2,k}+\ldots+\alpha_{j,j-1}\beta_{j-1,k}, \tag{7}$$

and the coefficients $\beta_{k+1,k}$, $\beta_{k+2,k}$, ... can be built up in successive steps.

(b) *Coefficients of the Polynomial of Best Fit.*

The polynomial $u_p(x)$ is written in the form

$$u_p(x)=a_pT_p(x)+a_{p-1}T_{p-1}(x)+\ldots+a_1T_1(x)+a_0, \tag{8}$$

where the coefficients a_j are to be chosen so that the expression

$$\Sigma v_p^2(x_i)=\Sigma\{y(x_i)-u_p(x_i)\}^2 \tag{9}$$

is a minimum. Differentiating the right-hand expression with respect to a_j to obtain the normal equations,

$$\Sigma\{y(x_i)-u_p(x_i)\}T_j(x_i)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

Then

$$\Sigma y(x)T_j(x)=a_j\Sigma T_j^2(x)$$

or

$$a_j=\Sigma y(x)T_j(x)/\Sigma T_j^2(x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (11)$$

The power moments M_j of the observations are defined in the usual way:

$$M_j=\Sigma x^j y(x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (12)$$

The orthogonal moments M'_j are defined by the equations

$$M'_j=\Sigma T_j(x)y(x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (13)$$

The equation for a_j may then be written in the form

$$a_j=M'_j/S_{jj}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (14)$$

Now $T_j(x)$ in (13) can be expanded by using either (2) or (6). We obtain respectively

$$\begin{aligned} a_j&=\Sigma y(x)\{x^j+\alpha_{j,j-1}T_{j-1}(x)+\ldots+\alpha_{j0}\}/S_{jj} \\ &=\{M_j+\alpha_{j,j-1}M'_{j-1}+\alpha_{j,j-2}M'_{j-2}+\ldots+\alpha_{j0}M'_0\}/S_{jj} \qquad . \qquad (15) \end{aligned}$$

and

$$\begin{aligned} a_j&=\Sigma y(x)\{x^j+\beta_{j,j-1}x^{j-1}+\ldots+\beta_{j0}\}/S_{jj} \\ &=\{M_j+\beta_{j,j-1}M_{j-1}+\beta_{j,j-2}M_{j-2}+\ldots+\beta_{j0}M_0\}/S_{jj}. \quad . \quad . \qquad (16) \end{aligned}$$

Either of the equations (15) and (16) can be used to calculate a_j.

The polynomial of best fit is most commonly required in the form of a power series in x,

$$u_p(x)=b_{pp}x^p+b_{p,p-1}x^{p-1}+\ldots+b_{p0}. \quad . \quad . \quad . \quad . \qquad (17)$$

Substitution of the expressions $T_j(x)$ given by (6) in (8), and collection of terms in x^j, leads to the equations

$$b_{pj}=a_j+\beta_{j+1,j}a_{j+1}+\beta_{j+2,j}a_{j+2}+\ldots+\beta_{pj}a_p. \quad . \quad . \quad . \qquad (18)$$

An alternative expression is found by substitution of the values $\beta_{j,k}$ from equation (7) and grouping of terms in $a_{j,k}$.

$$\begin{aligned} b_{pj}&=a_j+\alpha_{j+1,j}(a_{j+1}+\beta_{j+2,j+1}a_{j+2}+\ldots+\beta_{p,j+1}a_p) \\ &\qquad+\alpha_{j+2,j}(a_{j+2}+\beta_{j+3,j+2}a_{j+3}+\ldots+\beta_{p,j+2}a_p)+\ldots \\ &=a_j+\alpha_{j+1,j}b_{p,j+1}+\alpha_{j+2,j}b_{p,j+2}+\ldots+\alpha_{pj}b_{pp}. \quad . \quad . \quad . \quad . \qquad (19) \end{aligned}$$

Either of the equations (18) and (19) can be used to calculate the coefficients b_{pj}.

If the power-series expansion of $u_p(x_i)$ is substituted in (10), the normal equations become

$$\Sigma\{y(x_i)-(b_{p0}+b_{p1}x+\ldots+b_{pp}x^p)\}T_j(x)=0$$

for all j in the range 0 to p. Using (2), the set of normal equations reduces to

$$\Sigma\{y(x_i)-(b_{p0}+b_{p1}x+\ldots+b_{pp}x^p)\}x^j=0$$

or

$$b_{p0}S_{j0}+b_{p1}S_{j+1,0}+\ldots+b_{pp}S_{j+p,0}=M_j. \quad . \quad . \quad . \quad (20)$$

This is the form in which the normal equations are generally written. Substitution of the calculated coefficients b in the equations (20) (for $j=0$ and 1, say) provides a check on the arithmetical calculations.

(c) *Practical Methods of Calculation.*

There are two possible ways of carrying out the calculations; the first using the coefficients α, the second the coefficients β. In each system the α_{jk} and the S_{jj} are calculated. In the first system (the α-system) the coefficients α are used directly, while in the second system (the β-system) the coefficients β are determined from the α and then used in the remainder of the calculations. The appropriate equations for a and b are (15) and (19) in the α-system, (16) and (18) in the β-system.

The α-system is essentially the familiar Gauss-Doolittle method of solving the normal equations. Table I. shows a typical Doolittle scheme (based on that of Goulden (1939)), with the entries in our notation.

The principal disadvantage of the α-system is that the terms $\alpha_{kj}b_{pk}$ in the expression for b_{pj} are different for polynomials of different degree p, and have to be calculated separately for each degree. In the β-system the terms $\beta_{kj}a_j$ in the expression for b_{pj} are the same for polynomials of different degree.

The coefficients β_{jk} are needed for probable error calculations, and so if probable errors are to be calculated it would seem best to evaluate the β_{jk} and use the β-system. If a number of different sets of observations are to be fitted, the values x_i being the same for each set, the β_{jk} can be calculated once and for all. It is also possible to calculate the a_j by the α-system (as in the Doolittle scheme), then the values β_{jk}, and finally the b_{pj} as in the β-system. The β-system is used in the illustrative example in §4.

Aitken's scheme (1933–34 b) is essentially a variation of the β-system. The scheme is based on a method of solution of simultaneous equations developed by Aitken (1932). Since the scheme will hardly be understood without a knowledge of this method, a detailed account will not be given here. A feature of the scheme is that division is postponed throughout the table; for example, the entry corresponding to $S_{21}=S_{30}+\alpha_{10}S_{20}$ is written as $S_{00}S_{30}-S_{10}S_{20}=S_{00}S_{21}$ and the number S_{00} is entered in a divisor column. There are no non-terminating decimals in the table, but the later entries become very large.

In our notation the entries in the top right-hand line of each section of Aitken's scheme are (when divided by the divisor) the coefficients β_{jk}. Thus, although Aitken does not himself carry out probable error calculations, the required values β_{jk} are displayed in his table. As in our β-system, all the polynomials of lower degree are obtained in the process of calculating a polynomial of given degree.

TABLE I.

The Doolittle Scheme.

1.	S_{00}	S_{10}	S_{20}	S_{30}	M_0
2. Divide by $-S_{00}$		α_{10}	α_{20}	α_{30}	$-a_0$
3.		S_{20}	S_{30}	S_{40}	M_1
4. $\alpha_{10}\times(1)$		$\alpha_{10}S_{10}$	$\alpha_{10}S_{20}$	$\alpha_{10}S_{30}$	$\alpha_{10}M_0$
5. Add		S_{11}	S_{21}	S_{31}	M_1'
6. Divide by $-S_{11}$			α_{21}	α_{31}	$-a_1$
7.			S_{40}	S_{50}	M_2
8. $\alpha_{20}\times(1)$			$\alpha_{20}S_{20}$	$\alpha_{20}S_{30}$	$\alpha_{20}M_0$
9. $\alpha_{21}\times(5)$			$\alpha_{21}S_{21}$	$\alpha_{21}S_{31}$	$\alpha_{21}M_1'$
10. Add			S_{22}	S_{32}	M_2'
11. Divide by $-S_{22}$				α_{32}	$-a_2$
12.				S_{60}	M_3
13. $\alpha_{30}\times(1)$				$\alpha_{30}S_{30}$	$\alpha_{30}M_0$
14. $\alpha_{31}\times(5)$				$\alpha_{31}S_{31}$	$\alpha_{31}M_1'$
15. $\alpha_{32}\times(10)$				$\alpha_{32}S_{32}$	$\alpha_{32}M_2'$
16. Add				S_{33}	M_3'
17. Divide by $-S_{33}$					$-a_3$
$b_{33}=$					a_3
$b_{32}=$				$\alpha_{32}b_{33}$	a_2
$b_{31}=$			$\alpha_{21}b_{32}$	$\alpha_{31}b_{33}$	a_1
$b_{30}=$		$\alpha_{10}b_{31}$	$\alpha_{20}b_{32}$	$\alpha_{30}b_{33}$	a_0

(d) *Probable Errors.*

The sum of the squares of the residuals is given by the equation

$$\begin{aligned}\Sigma v_p^2&=\Sigma\{y(x_i)-\mu_p(x_i)\}^2\\&=\Sigma\{y(x_i)-a_0-a_1T_1(x_i)-\dots-a_pT_p(x_i)\}^2,\end{aligned}$$

and so, using (11),

$$\left.\begin{aligned}\Sigma v_p^2&=\Sigma y^2-S_{00}a_0^2-S_{11}a_1^2-\dots-S_{pp}a_p^2\\&=\Sigma y^2-M_0'a_0-M_1'a_1-\dots-M_p'a_p\\&=\Sigma y^2-M_0'^2/S_{00}-M_1'^2/S_{11}-\dots-M_p'^2/S_{pp}.\end{aligned}\right\}\qquad(21)$$

Σv_p^2 can also be obtained by calculating the various $u(x_i)$ directly. As a check on the arithmetical calculations it may be desirable to compute Σv_p^2 by both methods.

The probable error r_p of an observation $y(x_i)$ of unit weight is then given by the usual formula

$$r_p = 0{\cdot}6745\{\Sigma v_p^2/(n-p-1)\}^{\frac{1}{2}}. \quad \ldots \quad (22)$$

To find the probable errors of the coefficients of $u_p(x)$, these coefficients must be expressed as explicit functions of the observations. Thus

$$a_k = \Sigma\{\mathrm{T}_k(x_i)/\mathrm{S}_{kk}\}\,.\,y(x_i),$$

and hence

$$r^2[a_k] = \Sigma\left(\frac{\partial a_k}{\partial y(x_i)}\right)^2 r^2[y(x_i)]$$

$$= r_p^2/\mathrm{S}_{kk}. \quad \ldots \quad (23)$$

For the coefficients b_{pk},

$$r^2[b_{pk}] = \Sigma\left(\frac{\partial b_{pk}}{\partial y(x_i)}\right)^2 r^2[y(x_i)],$$

and, expanding b_{pk} as a function of the a_j using (18),

$$r^2[b_{pk}] = r_p^2\Sigma\{\mathrm{T}_k(x_i)/\mathrm{S}_{kk} + \beta_{k+1,\,k}\mathrm{T}_{k+1}(x_i)/\mathrm{S}_{k+1,\,k+1} + \ldots\}^2$$

$$= r_p^2\{1/\mathrm{S}_{kk} + \beta_{k+1,\,k}^2/\mathrm{S}_{k+1,\,k+1} + \ldots + \beta_{pk}^2/\mathrm{S}_{pp}\}, \quad \ldots \quad (24)$$

since the cross-products $\Sigma\mathrm{T}_j(x)\mathrm{T}_k(x)$ vanish.

Because the polynomials $\mathrm{T}_j x$ are orthogonal, the coefficients a_j behave as independently observed quantities in probable error calculations. For example, formula (24) can be written

$$r^2[b_{pk}] = \sum_{j=k}^{p}\left(\frac{\partial b_{pk}}{\partial a_j}\right)^2 r^2[a_j]. \quad \ldots \quad (25)$$

e) *Fitted Values.*

If the fitted values are required at various points x (which may or may not be members of the set x_i), they may be calculated by means of the power series (17). The probable errors of the fitted values at the points x are given by the equations

$$r^2[u_p(x)] = r^2[a_0] + \mathrm{T}_1^2(x)r^2[a_1] + \mathrm{T}_2^2(x)r^2[a_2] + \ldots$$

$$= r_p^2\{1/\mathrm{S}_{00} + \mathrm{T}_1^2(x)/\mathrm{S}_{11} + \mathrm{T}_2^2(x)/\mathrm{S}_{22} + \ldots\}$$

$$= r_p^2\{1/\mathrm{S}_{00} + (x+\beta_{10})^2/\mathrm{S}_{11} + (x^2+\beta_{21}x+\beta_{20})^2/\mathrm{S}_{22} + \ldots\}. \quad . \quad (26)$$

(f) *Changes of Origin and Scale.*

It is sometimes necessary to change the origin and scale of the variable x. Suppose it is desired to convert the polynomial $u_p(x)$ into a polynomial $u_p(\xi)$ in the new variable

$$\xi=(x+h)/m. \qquad (27)$$

Then

$$\begin{aligned} u_p(\xi) &= b_{p0}+b_{p1}(m\xi-h)+b_{p2}(m\xi-h)^2+\dots \\ &= c_{p0}+c_{p1}\xi+c_{p2}\xi^2+\dots, \qquad (28) \end{aligned}$$

where

$$c_{pj}=m^j\sum_{k=j}^{p}(-h)^{k-j}\binom{k}{j}b_{pk}. \qquad (29)$$

To calculate the probable errors it is necessary to express the coefficients as explicit functions of the a_k. Thus

$$c_{pj}=m^j\sum_{k=j}^{p}\gamma_{kj}a_k, \qquad (30)$$

where

$$\gamma_{kj}=\sum_{q=j}^{k}(-h)^{q-j}\binom{q}{j}\beta_{kq}, \qquad (31)$$

and so

$$r[c_{pj}]=m^j\left(\sum_{k=j}^{p}\gamma_{kj}^2/\mathrm{S}_{kk}\right)^{\frac{1}{2}}r_p. \qquad (32)$$

§ 3. TABLES OF SYMBOLS AND FORMULÆ.

(a) *Symbols.*

$y(x_i)$	observations.
$u_p(x)$	least squares polynomial of degree p which fits the $y(x_i)$.
$\mathrm{T}_j(x)$	Tchebycheff polynomial of degree j.
S_{j0}	Σx^j.
S_{jk}	$\Sigma x^j\mathrm{T}_k(x)$.
α_{jk}	coefficient of $\mathrm{T}_k(x)$ in orthogonal expansion of $\mathrm{T}_j(x)$.
β_{jk}	coefficient of x^k in power-series expansion of $\mathrm{T}_j(x)$.
M_j	$\Sigma x^j y(x)$, jth power moment of the observations.
M'_j	$\Sigma\mathrm{T}_j(x)y(x)$, jth orthogonal moment of the observations.
a_j	coefficient of $\mathrm{T}_j(x)$ in orthogonal expansion of $u_p(x)$.
b_{pj}	coefficient of x^j in power-series expansion of $u_p(x)$.
v_p	residual $y(x_i)-u_p(x_i)$.
r_p	probable error of an observation of unit weight.
$r[a_k]$	probable error of a_k.
$r[b_{kp}]$	probable error of b_{pk}.
$r[u_p(x)]$	probable error of fitted value $u_p(x)$.
ξ	$(x+h)/m$, change of variable.
c_{pj}	coefficient of ξ^j in power-series expansion of $u_p(\xi)$.

(b) *Formulæ.*

$$
\begin{aligned}
S_{j0} &= \Sigma x^j. \\
S_{jk} &= S_{j+k,0} + \alpha_{k,k-1} S_{j,k-1} + \alpha_{k,k-2} S_{j,k-2} + \ldots + \alpha_{k0} S_{j0}. \\
\alpha_{jk} &= -S_{jk}/S_{kk}. \\
\beta_{jk} &= \alpha_{jk} + \alpha_{j,k+1}\beta_{k+1,k} + \alpha_{j,k+2}\beta_{k+2,k} + \ldots + \alpha_{j,j-1}\beta_{j-1,k}. \\
M_j &= \Sigma y x^j. \\
M'_j &= M_j + \beta_{j,j-1} M_{j-1} + \beta_{j,j-2} M_{j-2} + \ldots + \beta_{j0} M_0. \\
&= M_j + \alpha_{j,j-1} M'_{j-1} + \alpha_{j,j-2} M'_{j-2} + \ldots + \alpha_{j0} M'_0. \\
a_j &= M'_j/S_{jj}. \\
b_{pj} &= a_j + \beta_{j+1,j} a_{j+1} + \beta_{j+2,j} a_{j+2} + \ldots + \beta_{pj} a_p. \\
&= a_j + \alpha_{j+1,j} b_{p,j+1} + \alpha_{j+2,j} b_{p,j+2} + \ldots + \alpha_{pj} b_{pp}. \\
\Sigma v_p^2 &= \Sigma y^2 - S_{00} a_0^2 - S_{11} a_1^2 - \ldots - S_{pp} a_p^2. \\
&= \Sigma y^2 - M'_0 a_0 - M'_1 a_1 - \ldots - M'_p a_p. \\
&= \Sigma y^2 - M_0'^2/S_{00} - M_1'^2/S_{11} - \ldots - M_p'^2/S_{pp}. \\
r_p &= 0{\cdot}6745\{\Sigma v_p^2/(n-p-1)\}^{\frac{1}{2}}. \\
r[b_{pj}] &= r_p(1/S_{jj} + \beta_{j+1,j}^2/S_{j+1,j+1} + \ldots + \beta_{pj}^2/S_{pp})^{\frac{1}{2}}. \\
u_p(x) &= b_{p0} + b_{p1} x + b_{p2} x^2 + \ldots + b_{pp} x^p. \\
r[u_p(x)] &= r_p\{1/S_{00} + T_1^2(x)/S_{11} + T_2^2(x)/S_{22} + \ldots + T_p^2(x)/S_{pp}\}^{\frac{1}{2}}. \\
T_j(x) &= x^j + \beta_{j,j-1} x^{j-1} + \beta_{j,j-2} x^{j-2} + \ldots + \beta_{j0}
\end{aligned}
$$

§ 4. MODEL FORMS AND ILLUSTRATIVE EXAMPLE.

In carrying out polynomial fitting, it is highly desirable to employ some form of tabulation for the various quantities to be calculated. " Model Forms " have been developed for the β-system, and these will be illustrated in this section by fitting a polynomial to the observations listed in Table II.

TABLE II.

x	y	x	y
−9·0	0·40	+0·8	2·30
−7·3	0·53	+2·8	2·85
−6·4	1·09	+3·5	3·62
−5·5	0·76	+4·8	4·50
−4·4	0·80	+6·1	6·40
−3·65	1·17	+6·25	5·08
−2·2	1·16	+7·2	6·85
−1·25	1·45	+7·9	8·10

A rough graph shows that a quadratic or cubic equation would probably provide a satisfactory fit for the points. The calculations will be carried through for the cubic: in the β-system the quadratic is obtained at the same time.

The first step is to calculate the sums S_{j0} of the powers x^j at the points of observation for $j=0$ to 6. This is best done by tabulating the powers of x in columns.

The second step is to calculate the sums M_j of the products $x^j y(x)$ at the points of observation. Again this is best done by tabulating the products in columns. The sum of the squares of the observations Σy^2 is calculated at the same time.

To save space the detailed calculations for these two steps are omitted.

Model Form A.

$\alpha_{jk} = -S_{jk}/S_{kk}$ $a_j = M'_j/S_{jj}$

S_{00}	16	S_{10} α_{10}	−0·35 +0·021875	S_{20} α_{20}	+478·8775 −29·929844	S_{30} α_{30}	−179·6086 +11·22554
α_{10}	+0·021875	S_{20} $\alpha_{10}S_{10}$	+478·8775 −0·0077	S_{30} $\alpha_{10}S_{20}$	−179·6086 +10·4754	S_{40} $\alpha_{10}S_{30}$	+22807·411 −3·929
$=\beta_{10}$	+0·021875	S_{11}	478·8698	S_{21} α_{21}	−169·1332 +0·3531925	S_{31} α_{31}	+22803·482 −47·61938
α_{20} $\alpha_{21}\beta_{10}$	−29·929844 +0·00773	α_{21}	+0·3531925	S_{40} $\alpha_{20}S_{20}$ $\alpha_{21}S_{21}$	+22807·411 −14332·729 −59·737	S_{50} $\alpha_{20}S_{30}$ $\alpha_{21}S_{31}$	−26553·43 +5375·66 +8054·02
β_{20}	−29·92211	$=\beta_{21}$	+0·3531925	S_{22}	8414·945	S_{32} α_{32}	−13123·75 +1·559576
α_{30} $\alpha_{31}\beta_{10}$ $\alpha_{32}\beta_{20}$	+11·22554 −1·04167 −46·66580	α_{31} $\alpha_{32}\beta_{21}$	−47·61938 +0·55083	α_{32}	+1·559576	S_{60} $\alpha_{30}S_{30}$ $\alpha_{31}S_{31}$ $\alpha_{32}S_{32}$	+1296991·1 −2016·2 −1085888 20467·5
β_{30}	−36·48193	β_{31}	−47·06855	$=\beta_{32}$	+1·559576	S_{33}	188619
$M_0=M'_0$	47·06	a_0	+2·941250			Σy^2	233·7294
M_1 $\beta_{10}M_0$	+197·4100 +1·0294			M'_0a_0	138·4152	Σv_0^2	95·3142
M'_1	+198·4394	a_1	+0·4143911	M'_1a_1	82·2315		
M_2 $\beta_{20}M_0$ $\beta_{21}M_1$	+1636·283 −1408·134 +69·724			M'_2a_2	10·5441	Σv_1^2 Σv_2^2	13·0827 2·5386
M'_2	+297·873	a_2	+0·0353981	M'_3a_3	1·0878		
M_3	+8909·704					Σv_3^2	1·4508
$\beta_{30}M_0$ $\beta_{31}M_1$ $\beta_{32}M_2$	−1716·840 −9291·802 +2551·908			r_3	$0{\cdot}674\sqrt{\left(\frac{v_3^2}{n-4}\right)}$ $=0{\cdot}234$		
M'_3	+452·970	a_3	+0·00240151				

Model Form A (*Cont.*)

$\alpha_{jk} = -S_{jk}/S_{kk}$ $a_j = M'_j/S_{jj}$

a_0 $= b_{00}$	$+2{\cdot}941250$						
$\beta_{10}a_1$ b_{10}	$+0{\cdot}009065$ $+2{\cdot}950315$	a_1 $= b_{11}$	$+0{\cdot}4143911$				
$\beta_{20}a_2$ b_{20}	$-1{\cdot}059186$ $+1{\cdot}891129$	$\beta_{21}a_2$ b_{21}	$+0{\cdot}0125023$ $+0{\cdot}4268934$	a_2 $= b_{22}$	$+0{\cdot}0353981$		
$\beta_{30}a_3$ b_{30}	$-0{\cdot}087612$ $+1{\cdot}803517$	$\beta_{31}a_3$ b_{31}	$-0{\cdot}113036$ $+0{\cdot}313857$	$\beta_{32}a_3$ b_{32}	$+0{\cdot}0037453$ $+0{\cdot}0391434$	a_3 $= b_{33}$	$+0{\cdot}00240151$
$1/S_{00}$ β^2_{10}/S_{11} β^2_{20}/S_{22} β^2_{30}/S_{33}	$6{\cdot}25 \times 10^{-2}$ 0 $10{\cdot}6$ $0{\cdot}71$	$1/S_{11}$ β^2_{21}/S_{22} β^2_{31}/S_{33}	$0{\cdot}21 \times 10^{-2}$ 0 $1{\cdot}18$	$1/S_{22}$ β^2_{32}/S_{33}	$1{\cdot}19 \times 10^{-4}$ $0{\cdot}13$	$1/S_{33}$	$5{\cdot}30 \times 10^{-6}$
$(r_b/r_3)^2$	$17{\cdot}6 \times 10^{-2}$	$(r_b/r_3)^2$	$1{\cdot}39 \times 10^{-2}$	$(r_b/r_3)^2$	$1{\cdot}32 \times 10^{-4}$	$(r_b/r_3)^2$	$5{\cdot}30 \times 10^{-6}$
$r[b_{30}]$	$9{\cdot}80 \times 10^{-2}$	$r[b_{31}]$	$2{\cdot}76 \times 10^{-2}$	$r[b_{32}]$	$2{\cdot}69 \times 10^{-3}$	$r[b_{33}]$	$5{\cdot}38 \times 10^{-4}$

(a) *Model Form A.*

The sums S_{j0} of the powers x^j are entered in the first section of Model Form A and the values S_{kk}, α_{jk} $(= -S_{jk}/S_{kk})$, and β_{jk} evaluated.

The power moments M_j are entered in the second section, and the orthogonal moments M'_j and the orthogonal coefficients a_j $(= M'_j/S_{jj})$ evaluated. On the right-hand side the quantity Σy^2 is entered, the sums of the squares of the residuals Σv_p^2 determined, and the probable error r_3 of an observation of unit weight calculated for the cubic.

Next the power-series coefficients b_{pj} are calculated. The Model Form gives the coefficients of the lower degree polynomials as well as those of the cubic.

Finally, the probable errors $r[b_{3j}]$ of the cubic coefficients b_{3j} are evaluated.

The coefficients b_{3j} were checked, using the normal equations

$$\left.\begin{aligned} S_{00}b_{30} + S_{10}b_{31} + S_{20}b_{32} + S_{30}b_{33} &= M_0, \\ S_{10}b_{30} + S_{20}b_{31} + S_{30}b_{32} + S_{40}b_{33} &= M_1, \end{aligned}\right\}$$

and as a further check the residuals were found by direct calculation of the fitted values. To save space, these check calculations are omitted.

The cubic of best fit is then

$$u_3(x) = (1{\cdot}804 \pm 0{\cdot}098) + (0{\cdot}314 \pm 0{\cdot}028)x + (0{\cdot}0391 \pm 0{\cdot}0027)x^2 + (0{\cdot}00240 \pm 0{\cdot}00054)x^3.$$

(b) *Model Form B.*

Suppose the fitted values $u(x)$ at the points

$$x = -20 \quad -15 \quad -10 \quad -5 \quad 0 \quad +5 \quad +10 \quad +15 \quad +20,$$

and the probable errors of these values, are required.

The quantities $T_j(x)$ at the required points are first calculated as in the upper section of Model Form B. These values are then used in the middle section to calculate the probable errors $r[u(x)]$. Finally the fitted values are calculated in the lower section.

All probable error calculations are carried out to slide-rule accuracy only.

Model Form B.

β	·0219			·353	−29·9			1·56	−47·1	−36·5	
x	β_{10}	$T_1(x)$	x^2	$\beta_{21}x$	β_{20}	$T_2(x)$	x^3	$\beta_{32}x^2$	$\beta_{31}x$	β_{30}	$T_3(x)$
−20	·0219	−20	400	−7	−29·9	363	−8000	624	+942	−36	−6480
−15		−15	225	−5		190	−3375	351	+706		−2360
−10		−10	100	−4		66	−1000	156	+471		−409
−5		−5	25	−1·8		−6·7	−125	39	+235		+113
0		0	0	0		−30	0	0	0		−36
+5		+5	25	+1·8		−3·1	+125	39	−235		−107
+10		+10	100	+4		74	+1000	156	−471		+649
+15		+15	225	+5		200	+3375	351	−706		+2980
+20		+20	400	+7		377	+8000	624	−942		+7640

S_{jj}	16	479	8410	189×10^3	r	0·234
x	$1/S_{00}$	T_1^2/S_{11}	T_2^2/S_{22}	T_3^2/S_{33}	$(r[u(x)]/r)^2$	$r[u(x)]$
−20	0·0625	0·84	15·6	222	238	3·6
−15		0·47	4·3	29·4	34·3	1·37
−10		0·208	0·518	0·89	1·68	0·30
−5		0·052	0·005	0·068	0·187	0·101
0		0	0·106	0·007	0·175	0·098
+5		0·052	0·001	0·061	0·177	0·098
+10		0·208	0·65	2·22	3·16	0·42
+15		0·47	4·7	47·2	52·4	1·70
+20		0·84	16·9	309	327	4·2

x	b_0	b_1x	b_2x^2	b_3x^3	$u_3(x)$
−20	1·804	6·278	15·64	19·2	−8·0 ±3·6
−15		4·709	8·80	8·10	−2·2 ±1·4
−10		3·139	3·91	2·40	+0·17 ±0·30
−5		1·570	0·978	0·300	+0·91 ±0·10
0		0	0	0	+1·80 ±0·10
+5					+4·65 ±0·10
+10					+11·25 ±0·42
+15					+23·4 ±1·7
+20					+42·9 ±4·2

§ 5. DISCUSSION OF ILLUSTRATIVE EXAMPLE.

(a) *Arithmetical Calculations.*

The calculations in § 4 were carried out on a hand-operated Brunsviga calculating machine. Seven-figure accuracy in multiplication and division was obtainable without trouble.

It will be observed that there is a loss of one significant figure in the calculations of S_{33}, a_2, a_3. In the present example the loss in accuracy is unimportant. If the scatter of the observations is small, it may be necessary to obtain the coefficients to a greater accuracy. This may be done by calculating more significant figures for the earlier quantities in Model Form A, particularly α_{10}, α_{20}, β_{20} and S_{11}. These can be calculated accurately without trouble, since the number of figures in S_{00}, S_{10}, S_{20} is usually small.

If the scatter of the residuals is small, it is desirable that the terms in the calculation of Σv^2 should be accurate to the same number of figures as Σy^2. It may be an advantage to use the alternative forms M_0^2/S_{00}, $M_1'^2/S_{11}$, etc. for these terms. If it is not possible to calculate the terms with sufficient accuracy, the individual residuals will have to be calculated.

Unless the person doing the fitting is an experienced computer, it is almost essential that each calculation should be checked before proceeding with the next. Whenever possible the check should be done in a different way. If desired, a check column can be introduced in the calculations of the quantities S, α, as in the Doolittle scheme. However, the check column does not provide a safeguard against small errors, since one or more of the values which go to make up the check column entry usually are considerably larger than the others.

In the fitting of curves of higher degree than the third, the arithmetic becomes extremely tiresome. Fortunately, in many applications, a quadratic or cubic provides a satisfactory fit. If it is necessary to fit a higher degree polynomial, the Model Forms may be extended to include the extra terms.

(b) *Choice of the Degree of the Polynomial.*

Birge (1947) has proposed a criterion using the values Σv^2 (as in Model Form A) for deciding on the degree of the most suitable polynomial. Increasing the degree from $j-1$ to j at first reduces Σv^2 by a large factor. Soon a value p is reached, such that for values of j greater than p the reduction in Σv^2 is small. The polynomial of degree p then provides the best fit for the observations.

In the illustrative example it would appear that the quadratic would give almost as good a fit as the cubic. However, if the individual residuals are examined, it will be found that the two observations at $x=6{\cdot}1$ and $x=6{\cdot}25$ contribute 1·10 to the value 1·45 of Σv^2. For the quadratic the contributions from these observations amount to 1·09 in a total of 2·54. In fact, any least squares polynomial of reasonable degree must pass more or less centrally between these points. It would be legitimate to subtract the contributions from these observations before comparing the values Σv^2. The contributions from the remaining observations amount to 1·45 and 0·35 for the quadratic and cubic respectively.

The cubic should then provide an appreciably better fit. Comparison of the graphs of the quadratic (dotted line) and cubic (solid line) in fig. 1 shows that this is indeed so.

Fig. 1.

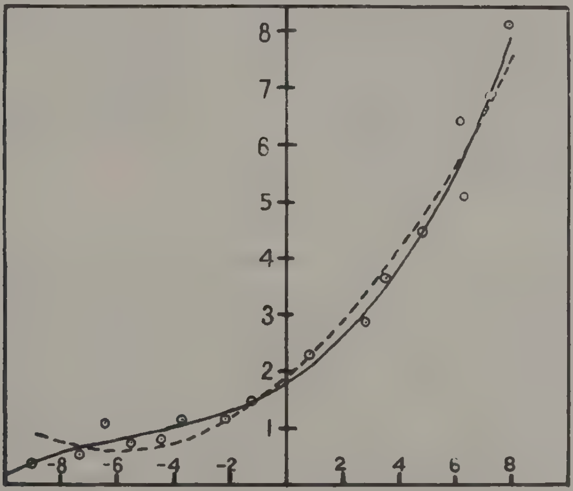


Fig. 1

References.

Aitken, A. C., 1932, *Proc. Edin. Math. Soc.*, **3** (2), 207; 1932–33, *Proc. Roy. Soc. Edin.*, **53**, 54; 1933–34 a, *Ibid.*, **54**, 1; 1933–34 b, *Ibid.*, **54**, 12.

Birge, R. T., 1947, *Rev. Mod. Phys.*, **19**, 298.

Birge, R. T., and **Shea**, J. D., 1927, *Univ. of Cal. Pub. in Math.*, **2**, 67.

Brunt, D., 1917, *The Combination of Observations* (Cambridge: University Press), Chap. 6.

Davis, H. T., 1935, *Tables of the Higher Mathematical Functions* (Bloomington, Indiana: Principia Press), **2**, 307–385.

Doolittle, M. H., 1878, *U.S. Coast and Geodetic Survey. Report*, Appendix 8, 115.

Fisher, R. A., 1946, *Statistical Methods for Research Workers*, Chap. 5, 10th ed. (Edinburgh: Oliver and Boyd).

Goulden, C. H., 1939, *Methods of Statistical Analysis* (New York: John Wiley & Sons), Chap. 14.

Isserlis, L., 1927, *Biometrika*, **19**, 87.

Kerawala, S. M., 1941, *Indian J. Phys.*, **15**, 241.

XI. *Emission of Protons and Alpha-Particles in High Energy Cosmic Ray Stars.*

By D. H. Perkins,
Imperial College of Science and Technology, London, S.W.7 *†.

[Received November 21, 1949.]

1. Summary.

This paper describes measurements of the energy, angular and frequency distributions of protons and alpha-particles in disintegrations of the silver and bromine nuclei of nuclear emulsions, at excitation energies comparable with and larger than the binding energy ($\sim$700 MeV.).

Only those charged particles with ionization greater than six times minimum have been considered (*i. e.* greater than that of a 50 MeV. proton). The star disintegrations were observed in Ilford C2 plates, 100 μ thick, exposed at Jungfraujoch (3457 m.).

2. Ratio of numbers of Alpha-particles to Protons from Stars.

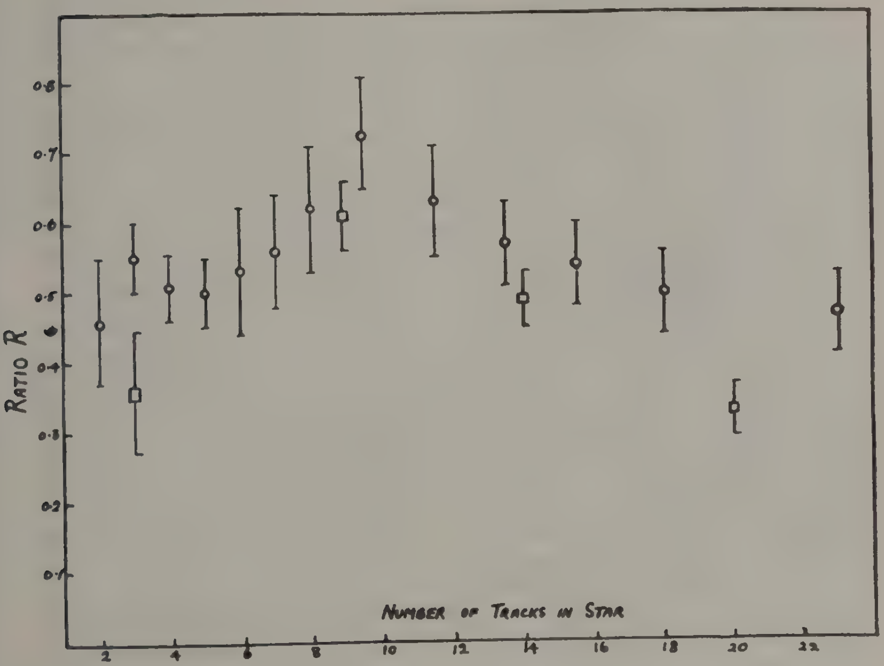
A quantity of interest in the theory of nuclear disintegrations is the ratio (α/p) of the numbers of alpha-particles to protons emitted at different excitation energies. Fig. 1 shows the result of measuring, for different prong number N, the ratio R of numbers of particles, charge $\geqslant 2$, to singly charged particles. It has already been shown (Harding, Lattimore and Perkins 1949) that the great majority of the singly charged particles ($\sim$85 per cent) consist of protons. In order to obtain the true α/p ratio in AgBr nuclei, however, two main corrections to R are required. For $N<6$, the α/p ratio in AgBr is less than R, because of the contribution of the light gelatine atoms in the emulsion (Perkins 1949 a). The latter have an α/p ratio of about unity. From the known proportion of stars of $N<6$ due to carbon, oxygen and nitrogen (Harding 1949 a), we thus obtain $(\alpha/p)_{\mathrm{AgBr}}$ for $N=3$ as indicated in the graph. Secondly, especially for high values of N, particles heavier than alpha-particles are frequently emitted from stars (Perkins 1949 b). The result of subtracting these from R is indicated for stars of 7–11 prongs ($\overline{N}=9$), 12–16 prongs ($\overline{N}=14$) and 17–31 prongs ($\overline{N}=20$). From the four true values of $(\alpha/p)_{\mathrm{AgBr}}$ thus obtained, it is clear that the α/p ratio is not a monotonic function of excitation energy. The initial increase is easily understandable, since as the temperature, and therefore kinetic energy of the emitted particles increases, the relative height of the potential barrier to alpha-particles and protons becomes less important. As N increases further, α/p appears to reach a broad maximum in the region $N=10$. A maximum in the ratio at just this excitation energy (~300 MeV) has been predicted by Le Couteur (1949) on the evaporation model. It is to be expected that the emission probability for alpha-particles will decrease when the nuclear temperature exceeds half the binding energy of a nucleon in an alpha-particle.

---

* Communicated by Sir George Thomson.

† Now at H. H. Wills Physical Laboratory, University of Bristol.

Fig. 1.



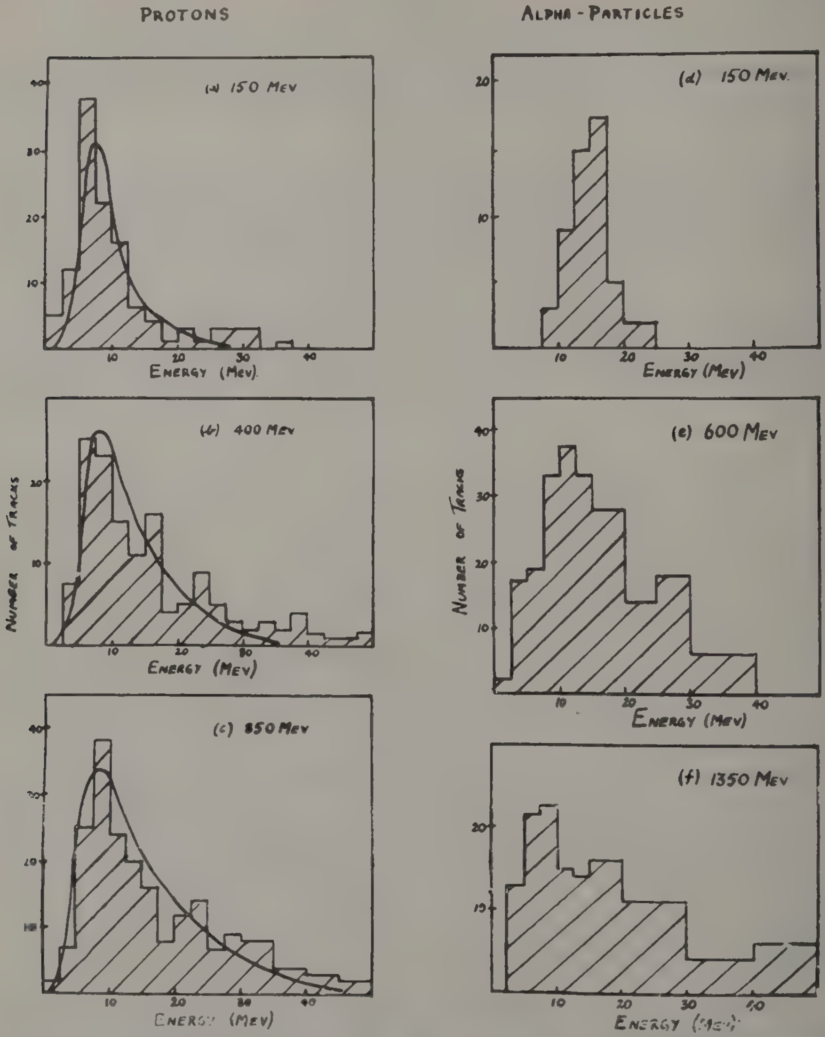
○ Ratio, R, of number of particles of charge greater than 2, to singly charged particles, for stars of different prong number. □ Corrected values of R, giving true alpha/proton emission ratio. Vertical lines indicate standard deviation.

3. Energy Distribution of Singly and Doubly Charged Particles.

(i) *Protons.*

The method of obtaining the energy distributions of the singly charged particles, predominantly protons, and alpha-particles has already been described (Harding, Lattimore and Perkins 1949). It was shown in this paper, hereafter referred to as HLP, that the energy distribution of the protons was consistent with the evaporation model of the nucleus. This model assumes that the energy transferred to the nucleus by the incident particle is distributed statistically among the nucleons, by successive collisions, before disintegration proceeds; thus any one nucleon has to undergo several collisions before it may acquire sufficient energy to escape from the nucleus (we exclude of course those nucleons "knocked on" directly by the star-producing particle, and contributing a high-energy "tail" to the Maxwellian distribution).

Fig. 2.



Energy distribution of protons and alpha-particles emitted from silver and bromine nuclei at different excitation energies. Protons: (*a*) 150 MeV.; (*b*) 400 MeV.; (*c*) 850 MeV. Full curves show the calculated evaporation distributions. Alpha-particles: (*d*) 150 MeV.; (*e*) 600 MeV.; (*f*) 1350 MeV.

The proton energy distributions, as given in HLP, are reproduced in fig. 2 (*abc*). Each histogram is taken from a group of stars with the mean excitation energies indicated. The latter were calculated by assuming the emission of that number of neutrons which would result in a stable residual nucleus. The curves were calculated assuming the free-particle model, with corrections. We see that the thermodynamical picture of the disintegration holds good for emission of protons (and presumably neutrons), even at excitation energies larger than the total nuclear binding energy ($\sim$700 MeV.). A more detailed analysis, also based on the Fermi-gas model, has recently been given by Le Couteur (1949) and by Fujimoto*.

(ii) *Alpha-particles.*

The energy distributions of the alpha-particles are also given in fig. 2 (*def*). In the distribution at 150 MeV. excitation energy (fig. 2 *d*) which corresponds to stars of small prong number, the contribution from the light emulsion nuclei has been subtracted by a method previously described (Perkins 1949a). Fig. 2 *d*, together with evidence described below, shows that the emission of alpha-particles in small stars may be treated as an evaporation process, most of the alpha-particles having energies greater than the barrier potential (10–12 MeV.). As the excitation energy increases, the distributions extend to higher energies, as do those of the protons, but we also obtain an increasing proportion of low-energy alpha-particles. In fact, the most probable energy of the alpha-particles *decreases* as the excitation energy increases. The reduction in effective barrier may perhaps be explained by a mean increase in nuclear radius due to surface oscillations, as suggested by Bagge (1943). Even so, it is not possible to understand why, if all the alpha-particles are released in a process of evaporation, there is a reduction in barrier for them but not for the protons. We must conclude that a considerable fraction, at least, of the alpha-particles are released in a different process. This will almost certainly be a preliminary process, when the nucleus is very " hot ", the reduction in barrier large, and thermal equilibrium not yet attained.

4. ZENITH-ANGLE DISTRIBUTION OF PROTONS AND ALPHA-PARTICLES.

A second difference in behaviour between protons and alpha-particles is revealed in their angular distributions with respect to the vertical.

If the stars are produced by primaries travelling downwards—in what follows we assume they are travelling vertically downwards—then the emitted particles should have a downward component of momentum. Let us assume that the alpha-particles, protons and neutrons leave the nucleus in an evaporation process, that is, the parent nucleus has taken up the whole momentum transfer from the primary before it disintegrates. Since, in producing a large star, the primary is likely to make several collisions in traversing the nucleus, the sideways components of momentum transfer tend to cancel out (alternatively, if there is only one very violent collision, several mesons may be produced and re-absorbed; all will be strongly directed downwards. This is equivalent to several collisions by the primary). We assume therefore that the net momentum transfer is vertically downwards, and furthermore that it is numerically equal to the energy transfer (*i. e.* the star is produced by a relativistic primary which also has relativistic energy after leaving the nucleus). All these assumptions seem reasonably well borne out by the observations of Harding (1949 b) on the smaller stars, and we want to investigate their validity in the larger ones.

---

* Y. Fujimoto, private communication.

Let u be the downward velocity of the nucleus as a result of the impact, $\bar{v}$ the average observed velocity of emission of an evaporating particle. We define $\bar{v}=\sqrt{(2\bar{E}/M)}$, where $\bar{E}$ is the average energy of a particle, M the mass. If $u \ll \bar{v}$, as is indeed always the case, it can be shown that, assuming a Maxwellian energy distribution (HLP), the downward excess of particles should constitute a fraction $\sqrt{(\pi/2)}\,.\,u/\bar{v}$ of the total, since the angular distribution of the particles in the coordinate system of the nucleus will be isotropic. u, as calculated from the mass of the nucleus and the energy transfer, can therefore be compared with the u obtained from the observed fractional downward excess of particles.

In the accompanying table, it can be seen that, if we take protons, neutrons and alpha-particles together, there is reasonably good agreement between the calculated and observed values of u (we assume the fractional downward excess of the neutrons, and their mean emission velocity, equal to that of the protons). In the group of smaller stars (7–11 prongs), with excitation energies well below the total binding energy, the ratio of downward velocities u_α/u_p, of alpha-particles to protons does not differ significantly from unity, in agreement with the hypothesis of evaporation. At higher energies, however, the downward excess of alpha-particles becomes extremely small and that of the protons correspondingly larger.

Table I.

Number of tracks in star	17–31	12–16	7–11
Estimated excitation energy (MeV.) =momentum transfer (MeV./c)	1350	750	420
Downward velocity of nucleus $=u/c$	*0·013*	*0·008*	*0·0045*
Fractional downward excess of—			
protons	(98±24)/600	(53±23)/513	(21±16)/279
alpha-particles	(6±18)/322	(12±17)/296	(18±13)/168
Average downward velocity of—			
(alphas+protons+neutrons) ..	*0·013±0·004*	*0·008±0·004*	*0·007±0·004*
protons u_p/c	0·025±0·006	0·014±0·006	0·009±0·007
alphas u_α/c	0·002±0·006	0·003±0·004	0·006±0·004
Ratio of downward velocities u_α/u_p	0·08 ±0·24	0·21 ±0·28	0·68 ±0·74

It is not of course possible to say if the process responsible for this behaviour, which appears to set in at excitation energies above 400 MeV., is also that which produces the reduction in effective barrier described above. A possible explanation, put forward by HLP, is as follows. Assume that, in a preliminary process, there is a " boiling off " of alpha-particles following intense localized " heating " in the top of the nucleus. Those travelling upwards escape clean away from the nucleus, whilst those travelling downwards are stopped by collisions and lead to a general warming up of the entire nucleus, succeeded by evaporation. The downward excess of alpha-particles which subsequently evaporate together with the protons may thus be cancelled out, and the whole nucleus will receive an additional downward momentum from the alphas travelling upwards in the preliminary process, thus accounting, in some degree at least, for the large downward excess of protons. Poor statistics do not allow us to investigate such processes in detail.

A second hypothesis, which is supported by experimental evidence, is that heavy fragments are emitted at high excitation energies, and that these are usually unstable and disintegrate spontaneously into alpha-particles. The latter then have to overcome a potential barrier much lower than that of the parent nucleus. Experimental evidence on the distribution of such fragments (Perkins 1949 b) is at present not sufficient to throw light on any resulting anomaly in the zenith-angle distribution of the alpha-particles.

It is to be pointed out that the observations above were made on those protons of energies below 50 MeV. However, these would include a few not produced in the evaporation process, but knocked-on directly by the primary particle. Suppose these knock-ons constitute a fraction f of all the protons. Since they should all travel downwards, they contribute approximately an additional fractional downward excess f, if $f \ll 1$. f is difficult to estimate, but it can be seen from the histograms of fig. 1 that it is not likely to exceed 0·1 in the large stars (*i. e.* less than half the total observed excess for protons). This can to a large extent account for the increased downward excess in such stars.

It can be concluded that both the energy and zenith-angle distributions of protons, and presumably neutrons, can be explained very satisfactorily on the evaporation picture, even at excitation energies considerably in excess of the total nuclear binding energy; but that the corresponding distributions for alpha-particles, in stars involving more than 400 MeV. energy, are inconsistent with such a picture. The elucidation of the mechanism of alpha-particle emission awaits further experimental evidence.

ACKNOWLEDGMENTS.

I should like to express my thanks to Dr. E. P. George and his staff, of Birkbeck College, London. About a third of all the stars represented in the above analysis were taken from plates kindly loaned by Dr. George, and without his generous cooperation these measurements would not have been possible.

REFERENCES.

BAGGE, E., 1943, *Physik. Zeit.*, **44**, 461.
LE COUTEUR, K. J., 1949 (in publication).
HARDING, J. B., 1949 a, *Nature*, **163**, 440.
HARDING, J. B., 1949 b, *Phil. Mag.*, **40**, 530.
HARDING, LATTIMORE and PERKINS, 1949, *Proc. Roy. Soc.* A, **196**, 325.
PERKINS, D. H., 1949 a, *Phil. Mag.*, **40**, 601.
PERKINS, D. H., 1949 b, Como Conference, September.

XII. *On the Fine Structure near the FeK-absorption Edges in the Various Oxides of Fe.*

By D. Coster and S. Kiestra,
Natuurkundig Laboratorium der Rijks-Universiteit, Groningen,
The Netherlands*.

[Received November 1, 1949.]

Summary.

After a review of the theoretical interpretation of the fine structure in X-ray absorption edges, new measurements on the K-edges of different compounds of iron at low dispersion are described. From these in conjunction with our earlier measurements at high dispersion (Coster and Kiestra 1948)† the following conclusions may be drawn with respect to these compounds.

(*a*). In a region up to about 70 eV. from the edge the behaviour of the absorption coefficient is mainly a property of the atom in question.

(*b*). In a region between 70 and 175 eV. the fine structure is determined by the immediate surroundings of the atom.

(*c*). In the region above 175 eV. the fine structure depends essentially on the crystal lattice.

This is in accordance with the notion that as the energy of the ejected electron increases it penetrates further into the lattice.

§1. Theoretical Foundations.

If the absorption coefficient of X-rays is considered in its dependence upon the frequency of the radiation, it may in general be stated that the absorption decreases with increasing frequency. Every time, however, that a critical frequency is surpassed, the absorption suddenly increases. This is caused by the fact that at these frequencies a new possibility of absorption arises, accompanying the removal of, for instance, a K- or L-electron from the atom. The corresponding phenomenon is called the K- or L-edge. On the short wavelength side of these edges the variation in the absorption is not monotonic, but fluctuates over a distance of several hundred volts. The difference between successive maxima and minima of the absorption coefficient may easily amount to 10–20 per cent. Of this phenomenon, which is called the fine structure of X-ray absorption edges, a theory based on wave mechanics was given by Kronig (I., 1931 ; II., 1932) for the case where the absorbing element forms part of a crystalline solid. According to this theory the material waves, representing an electron which by absorption of an X-ray quantum is removed from an inner shell of an atom and traverses the crystal lattice in a certain direction with a certain energy, suffer diffraction at the crystal lattice. The Bragg reflections, occurring at certain wavelengths and preventing the electron from entering the crystal (forbidden energy zones), even when averaged over all directions of ejection, still leave a fluctuation in the absorption probability as function of the wavelength of the electron or of the corresponding X-ray frequency.

---

* Communicated by the Authors.

† In the following part we will indicate this article as I,

Already in his first paper Kronig was able to show that several peculiarities of the fine structure of X-ray absorption edges could be explained on the basis of his theory. It is for instance evident that metals, crystallizing in the same lattice type, should have the same type of fine structure. It could indeed be shown experimentally (Coster and Veldkamp 1932) that for ejected electrons of the same sort (s- or p- etc. electrons) and for atomic numbers not too far apart this was the case. Furthermore, the distances of the maxima and minima of absorption from a suitable zero point corresponding to the average potential in the crystal appeared to be inversely proportional to the square of the lattice constant, as was predicted by the theory. If the crystal structure is changed by using an alloy instead of the pure element, the fine structure changes in the expected way (Coster and Klamer 1934, I.), the same being true in those cases where the crystal type is altered by heating (Coster and Levi 1939).

The fine structure of X-ray absorption edges is also observed when the absorbing element forms part of a polyatomic molecule (Coster and Klamer 1934, II.). For this case too a theory was given by Kronig (1932 III.). By absorption of an X-ray quantum an electron of an inner shell is removed with an excess of energy from the atom. The corresponding material waves are scattered by the other atoms of the molecule, and the resultant wave built up from the primary and the scattered waves determines the probability of absorption of the different X-ray frequencies. At first sight this method of calculation of fine structure might seem in principle to be different from that used in the case of crystalline solids ; on closer inspection, however, it is clear that also in the last case, instead of regarding the Bragg reflection of the material waves at the crystallographic planes, we may equally well determine the resultant wave built up by the primary wave and the secondary ones scattered by the surrounding atoms. In this process the atoms at short distance play a larger part than those farther removed. Coster (1935) was able to show that metals crystallizing in a cubic close-packed lattice have the same type of fine structure as those crystallizing in a hexagonal close-packed lattice, as long as attention is paid only to the relative *positions* of the successive maxima and minima of absorption, while some differences are noted when also their *form* is taken into account. This was associated with the fact that in both types of crystal lattices the immediate surroundings of an atom are formed by the same number of atoms, whereas differences in the arrangement of surrounding atoms are only observable for larger distances.

Kronig's theory treats the electron liberated from the atom as almost free, so that the forces exercised by the lattice may be regarded as small perturbations. This implies that the rules stated above are valid only at some distance from the zero potential. In the case of metal crystals this distance appeared to be more than 70 volts from the edge. Moreover, on the experimental side, the fine structure, as it is usually observed, arises by the absorption of non-polarized X-rays in a polycrystalline substance and consequently the fluctuations in the absorption coefficient which still remain represent an effect, averaged over all directions of ejection in the crystal. We thus understand that the fine structure of absorption edges is a rather rough phenomenon, and it may be regarded as a fortunate circumstance that in verifying experimentally the above-mentioned rules use was made in the present laboratory of X-ray spectrographs with not very high dispersion. Indeed, it is not difficult to ascertain that by using too large a dispersion we should easily run the risk of overlooking the general features mentioned.

If, however, it is desired to study the phenomena arising within a short distance from the main edge, the dispersion formerly used in this laboratory is no longer sufficient. A larger dispersion may be obtained firstly by the bent crystal method used by Cauchois (1948) and co-workers and Sanner (1941). Some American investigators (Barnes 1933, Beeman and Friedman 1939, Bearden and Snyder 1941), using the double crystal spectrometer, got very interesting results not obtainable with a smaller dispersion. They investigated a region within about 50 volts from the main edge, about which Kronig's theory does not give any information. Indeed, to explain the fine structure of X-ray absorption edges, in general, one has to bear in mind that in the corresponding transition an electron from an inner shell is removed to one of the unoccupied levels in the lattice. To calculate the corresponding absorption it is necessary to determine the density of states of these unoccupied levels. This can be done in two different ways :—

(1*e*) By starting from the wave function of a free electron.

(2*e*) By starting in zero approximation from the wave function of a free atom.

As has already been stated, the first procedure has been employed by Kronig, who calculated the perturbation caused by a periodic potential that serves to approximate the lattice field. It can only be applied if the energy of the electron in the lattice is large compared with the variations in the potential.

The second method has been developed by Wigner and Seitz (1933) and Slater (1934). It has been applied by Krutter (1935) to calculate the band structure in the case of Cu for the occupied levels with highest energy and the unoccupied levels to a distance of 20 volts from the Fermi-limit.

The K-absorption edge of Cu, taken by Beeman and Friedman (1939) is in good agreement with the theory of Krutter. The agreement is even better than could be anticipated from the approximations used.

In recent times it has been shown that by using the photographic method with the single crystal tube spectrograph the same dispersion and the same resolving power may be obtained as with the double crystal spectrometer (Coster and de Lang 1947). As the dispersion and resolving power of the spectrograph in question depend upon its dimensions, it is necessary to enlarge these. By constructing a spectrograph with a radius about five times as large as that formerly used in this laboratory, we (Coster and Kiestra 1948) were actually able to obtain results comparable in quality to those of the American authors just mentioned. At the same time, by applying some new devices we succeeded in compensating for the loss of energy caused by the increased dimensions.

While hitherto attention has been directed principally to metals, especially those of the transition periods, it seemed important to carry out an analogous investigation for compounds. There the situation is more involved due to the presence of several kinds of atoms, the greater complication of the lattice structure and the fact that a given sort of atom may have different valencies and even different surroundings in the same lattice. Measurements at high dispersion have already been published by Yoshida (1941, compounds of Fe, dispersion not very great), Sanner (1941, compounds of many transition metals), Beeman and Bearden (1942, aqueous solution of salts of Ni, Cu and Zn), Coster and Kiestra (1948, oxides of Fe and Mn) and Hanson and Beeman (1949, compounds of Mn).

In order to bring out also in the case of compounds the typical difference in the behaviour of the fine structure close to and far from the edge, we present in this publication results on the oxides of Fe taken at low dispersion with the purpose of comparing them with our earlier measurements on the same compounds at high dispersion.

§2. EXPERIMENTAL METHOD.

For these experiments the same spectrograph was used as for the investigations at high dispersion in I. The tube between the X-ray tube and the crystal and the tube containing the photographic plate holder were replaced by others, the distance focus-crystal and also crystal-photographic plate being reduced from 115 to 32·5 cm. in each case. In this way the dispersion decreased in the ratio 3·5 while the region investigated increased in the same ratio, *i. e.* to 400 eV. from the absorption edge.

Instead of tungsten molybdenium was now used as target to avoid the tungsten $L\eta$ emission line at 1675 X.U. Besides the tungsten hot wire was replaced by an oxide-coated platinum cathode. Care was taken that the $\mathrm{Ni}K\alpha_{12}$ lines were absent as much as possible so that also the fine structure in the region of 1660 X.U. could be determined.

The analysing crystal was a calcite crystal that remained fixed during the exposures. In the beginning the crystal had been carefully ground, polished and etched. A time of about 20 hours was then necessary for one exposure. As the resolving power of the spectrograph is relatively small at these dimensions, it became evident that the crystal does not need to have a high resolving power either. Therefore during the last exposures a crystal was used that had been ground, polished, but not etched, giving the same results in about five hours. This reduction in the time of exposure was very favourable, the intensity of the scattered radiation being diminished by a factor 4, compared with the former exposures.

Agfa Blau Hart plates were used. These are not as sensitive as Ilford Zenith plates but they have a finer grain. On the other hand always at least two exposures of the same substance were made and only those peculiarities were taken for granted which were present on two (or more than two) plates.

From the photographic plates photometer curves were taken with a Moll-photometer, using a beam of light of a height of 3 cm., which is the height of the photographed spectrum. In this way the grains of the plates do not spoil the curves very much.

In I. all the photometer curves were changed into intensity curves with the aid of a blackening scale for each plate. As can be seen from the figs. 5 and 6 in I. on the one hand and figs. 7 and 8 on the other, the main form of the curves as a whole is practically the same. For that reason in the present article only the photometer curves are given. The most interesting feature of these curves is the position and form of the maxima and minima of absorption.

§3. RESULTS.

In order to bring out the influence of the dispersion on the aspect of the fine structure we reproduce in figs. 1 and 2 the results for the K-edge found for Fe and published already elsewhere*. Furthermore, we show in figs. 3 and 4 the analogous photometer curves as found for various oxides of Fe, while fig. 5 gives a comparison at low dispersion of FeF_2, Fe_3O^4 and FeS.

As appears from figs. 1 and 2, the diminution in the dispersion causes a reduction in resolving power. In fact the minima α and β in fig. 1 are no longer separated in fig. 2. In addition the structure in the edge has become entirely invisible.

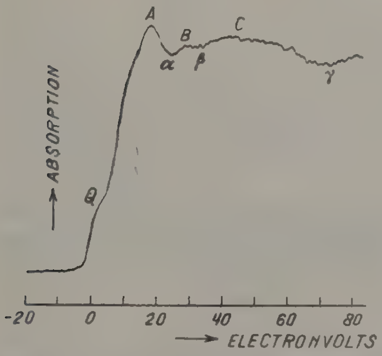
Fig. 1, embracing only a region of about 70 volts, is identical with the curve obtained for metallic iron by Beeman and Friedman (1939) with the double crystal spectrometer and entirely of the same type as the curves obtained for metallic manganese by Coster and Kiestra (1948) and by Hanson and Beeman (1949). An interpretation of this curve with which we can fully agree, has been given for the first time by Beeman and Friedman on the basis of the theoretical considerations of Krutter (1935) for Cu, already mentioned in the introduction.

---

* We call attention to the fact that just as in I. we deviate from the usage in the earlier literature by taking increasing absorption in the upward direction and increasing energy to the right. In addition for Fe the designation of the maxima differs from that of Coster and Veldkamp, an additional maximum having been discovered at the higher dispersion.

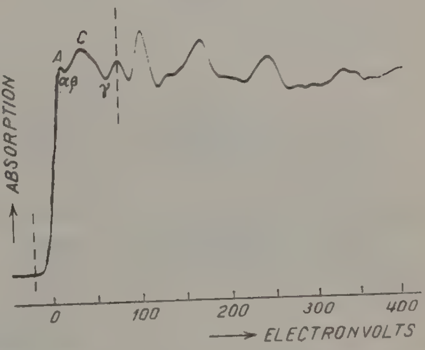
Fig. 2, on the other hand, at a distance of more than 70 volts from the main edge to a distance of about 400 volts, shows the typical fine structure of the body-centred cubic lattice of which the form is in excellent agreement with the theory of Kronig. How strongly this fine structure is characteristic for this crystal type has moreover been shown by Coster and Veldkamp (1932) and Coster and Klamer (1934, I.).

Fig. 1.



Photometer curve at ½ natural size of Fe K-absorption edge of metallic iron at high dispersion.

Fig. 2.

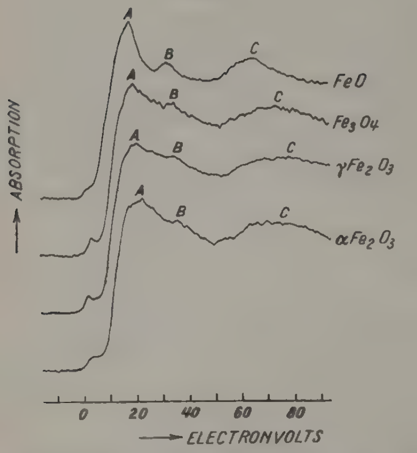


Photometer curve at ½ natural size of Fe K-absorption edge of metallic iron at low dispersion. The region corresponding to the whole of fig. 1 is denoted by dotted lines.

Comparison of figs. 3 and 4 shows that here, too, at low dispersion the maxima α and β are only separated for FeO. The initial absorption, clearly apparent at high dispersion as a small maximum to the left of the main edge, is practically unresolved at low dispersion.

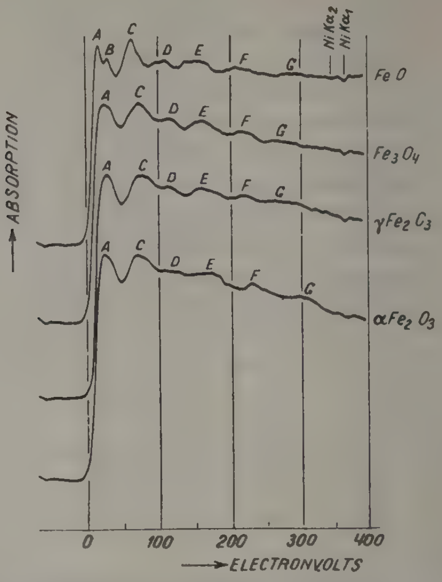
The most striking feature of fig. 4 is that the form of the fine structure in the region between 70 and 175 eV. is almost identical for all the oxides : two minima with one maximum (D) in between. Fig. 5 shows on the other hand that for FeS and FeF_2 in this region the structure is quite different. From this we conclude that the fine structure here is mainly due to the O-ions surrounding the Fe-ion. An additional argument is that the extrema for the ferrous ion, which is larger, lie closer to the edge than those for the smaller ferric ion. The fine structure in this case may be compared with that of polyatomic molecules in gases. The result is also reminiscent of that obtained by Coster (1935) in the case of cubic face-centered and hexagonal lattices already mentioned in § 1.

Fig. 3.



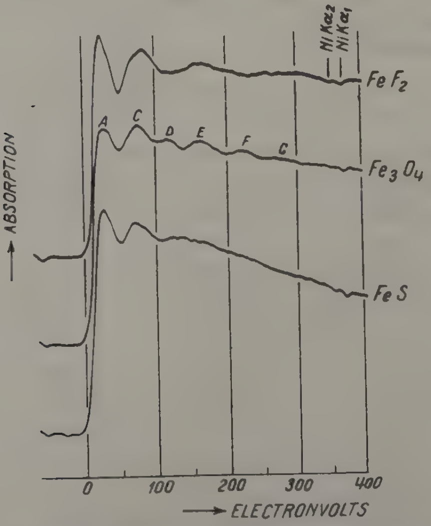
Photometer curves at $\frac{2}{5}$ natural size of Fe K-absorption edges of the iron oxides at high dispersion.

Fig. 4.



Photometer curves at $\frac{2}{5}$ natural size of Fe K-absorption edges of the iron oxides at low dispersion.

Fig. 5.



Photometer curves at $\frac{2}{5}$ natural size of Fe K-absorption edges of three iron compounds at low dispersion.

Regarding the fine structure for the oxides above 175 eV. we see that the extrema lie in different places for α Fe_2O_3 and γ Fe_2O_3, while γ Fe_2O_3 and Fe_3O_4 behave in an identical way. Now it is known that the α- and γ-modifications have different crystal structures, while γ Fe_2O_3 and Fe_3O_4 both are spinels with almost the same lattice constant. From this one concludes that the fine structure above 175 eV. is principally determined by the lattice type of these compounds. For crystals of metals the same is true above 70 eV.

Summarizing, we may say that as the energy of the ejected electron is increased the region of the crystal that plays a rôle in determining the fine structure is extended more and more.

The writers are greatly indebted to Professor R. Kronig for his advice in the completion of this report.

References.

BARNES, A. H., 1933, *Phys. Rev.*, **44**, 141.

BEARDEN, J. A., and SNYDER, T. M., 1941, *Phys. Rev.*, **59**, 162.

BEEMAN, W. W., and BEARDEN, J. A., 1942, *Phys. Rev.*, **61**, 455.

BEEMAN, W. W., and FRIEDMAN, H., 1939, *Phys. Rev.*, **56**, 392.

CAUCHOIS, Y., 1948, *Les spectres de rayons X et la structure électronique de la matière*.

COSTER, D., 1935, *Physica*, **2**, 606.

COSTER, D., and KIESTRA, S., 1948, *Physica*, **14**, 175.

COSTER, D., and KLAMER, G. H., 1934, I., *Physica*, **1**, 145 ; II., *Physica*, **1**, 879.

COSTER, D., and DE LANG, H., 1947, *Physica*, **13**, 379.

COSTER, D., and LEVI, H., 1939, *Physica*, **6**, 44.

COSTER, D., and VELDKAMP, J., 1932, *Zs. f. Phys.*, **74**, 191.

HANSON, H., and BEEMAN, W. W., 1949, *Phys. Rev.*, **76**, 118.

KRONIG, R., I., 1931, *Zs. f. Phys.*, **70**, 317 ; II., 1932, *Zs. f. Phys.*, **75**, 191 ; III., *Ibid.*, 468.

KRUTTER, H., 1935, *Phys. Rev.*, **48**, 664.

SANNER, V. H., 1941, thesis Uppsala.

SLATER, J. C., 1934, *Phys. Rev.*, **45**, 794.

WIGNER, E., and SEITZ, F., 1933, *Phys. Rev.*, **43**, 804.

YOSHIDA, S., 1941, *Sci. Papers Inst. Phys. Chem. Res.*, Tokyo, **38**, 272.

XIII. *Gravitational Fields of Spherical Symmetry and Weyl's Conformal Curvature Tensor.*

By V. V. Narlikar and K. P. Singh,
Benares Hindu University, India *.

[Received June 7, 1947.]

Abstract.

Solutions of the gravitational field equations of relativity are considered, so that the metric is non-static and has spherical symmetry, satisfying at the same time the additional equations, $C^h{}_{ijk,\,h}=0$, where the tensor C_{hijk} is Weyl's conformal curvature tensor. The additional equations always hold good, whatever be the metric, when $R_{ij}=0$, but when $R_{ij}\neq 0$ it is shown that the only spherically symmetrical solution of physical significance is what is known as Schwarzschild's internal solution. For the most general gravitational metric of spherical symmetry considered here, the condition $C_{hijk}=0$ reduces to one equation only, which is one of those following from $C^h{}_{ijk,\,h}=0$; and thus the corresponding space-time is conformal to flat space-time even when the latter condition alone is satisfied.

1. Introduction.

A gravitational field is distinguished from flat space-time by the conditions,

$$\left.\begin{array}{ll}\text{(i.)} & R_{hijk}=0, \text{ for flat space-time,}\\ \text{(ii.)} & R_{hijk}\neq 0, \text{ for a gravitational field,}\end{array}\right\}, \quad . \quad . \quad (1)$$

where R_{hijk} is the Riemann–Chrystoffel curvature tensor. When (ii.) is satisfied the distribution of matter and motion follows from the contracted tensor R_{ij}. Weyl's conformal curvature tensor C_{hijk} connects the contracted and original tensors by the equation (Eisenhart 1926):

$$C_{hijk}=R_{hijk}+\tfrac{1}{2}(g_{jh}R_{ik}-g_{hk}R_{ij}+g_{ik}R_{jh}-g_{ij}R_{hk})+\frac{R}{6}(g_{hk}g_{ij}-g_{hj}g_{ik}). \quad . \quad (2)$$

It is well-known that

$$C_{hijk}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

is a necessary and sufficient condition for the space-time to be conformal to flat space-time. It can be immediately verified that the contracted tensor, C_{ij}, vanishes identically:

$$g^{hk}C_{hijk}\equiv C_{ij}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

---

* Communicated by the Authors.

Thus the very structures of the three tensors give ten independent components for C_{hijk}, ten for R_{ij} and 20 for R_{hijk}. Moreover, the Bianchi identity, viz.,

$$R^h{}_{ijk,l}+R^h{}_{ikl,j}+R^h{}_{ilj,k}=0, \qquad \ldots \ldots \ldots \quad (5)$$

leads to

$$C^h{}_{ilj,h}=\tfrac{1}{2}R_{ilj}, \qquad \ldots \ldots \ldots \quad (6)$$

where (Eisenhart 1926)

$$R_{ilj}=R_{il,j}-R_{ij,l}+\tfrac{1}{6}(g_{ij}R_{,l}-g_{il}R_{,j}), \qquad \ldots \ldots \quad (7)$$

in the usual notation. It is clear from (6) that in the unoccupied portions of a gravitational field where

$$R_{ij}=0, \qquad \ldots \ldots \ldots \quad (8)$$

we have also

$$C^h{}_{ilj,h}=0. \qquad \ldots \ldots \ldots \quad (9)$$

(9) is also true when the space-time is conformal to flat space-time, that is, when (3) holds good, whatever may be the distribution of matter and motion.

Apart from the considerations outlined here little is known of the gravitational significance of the tensor C_{hijk}. The following is an attempt to investigate it further in relation to distributions of spherical symmetry for which the most general metric may be taken as

$$ds^2=-e^{\lambda}dr^2-r^2(d\theta^2+\sin^2\theta\, d\phi^2)+e^{\nu}\, dt^2, \qquad \ldots \quad (10)$$

where $\lambda=\lambda\,(r, t)$, $\nu=\nu\,(r, t)$. While (3) implies (9) in the general case it is found that, for (10), the latter implies the former.

2. The Condition, $C_{hijk}=0$.

For the metric (10) the components of R_{ij} which are not identically zero are

$$R_{11},\ R_{22},\ R_{33},\ R_{44},\ R_{14},\ R_{41}$$

with

$$R_{33}=R_{22}\sin^2\theta, \quad R_{14}=R_{41}. \qquad \ldots \ldots \quad (11)$$

They condition (3) implies ten equations, many of which are identically satisfied, the rest giving only one independent equation, viz.,

$$(1-e^{-\lambda})+\tfrac{1}{4}re^{-\lambda}(\nu'-\lambda')^2+r^2e^{-\nu}(\tfrac{1}{2}\ddot{\lambda}-\tfrac{1}{4}\dot{\lambda}\dot{\nu}+\tfrac{1}{4}\dot{\lambda}^2)-r^2e^{-\lambda}(\tfrac{1}{2}\nu''^2+\tfrac{1}{4}\nu'^2-\tfrac{1}{4}\lambda'\nu')=0 \qquad \ldots \ldots \ldots \quad (12)$$

Here, and in what follows, overhead dashes and dots represent differentiations with respect to r and t respectively. For the metric (10), the field equations of gravitation (Tolman 1934) give

$$8\pi T_1^1 = -e^{-\lambda}\left(\frac{\nu'}{r}+\frac{1}{r^2}\right)+\frac{1}{r^2}, \qquad (13)$$

$$8\pi T_2^2 = 8\pi T_3^3 = -e^{-\lambda}\left(\frac{\nu''}{2}-\frac{\lambda'\nu'}{4}+\frac{\nu'^2}{4}+\frac{\nu'-\lambda'}{2r}\right)+e^{-\lambda}\left(\frac{\ddot{\lambda}}{2}+\frac{\dot{\lambda}^2}{4}-\frac{\dot{\lambda}\dot{\nu}}{4}\right), \qquad (14)$$

$$8\pi T_4^4 = e^{-\lambda}\left(\frac{\lambda'}{r}-\frac{1}{r^2}\right)+\frac{1}{r^2}, \qquad (15)$$

$$8\pi T_1^4\, e^{\nu} = -8\pi T_4^1 e^{\lambda} = \frac{\dot{\lambda}}{r}. \qquad (16)$$

Hence (12) may be expressed as

$$3(1-e^{-\lambda})+8\pi r^2(T_2^2-T_1^1-T_4^4)=0. \qquad (17)$$

For a perfect fluid described by the metric (10),

$$T_1^1+T_4^4=\rho-p, \qquad T_2^2=T_3^3=-p, \qquad (18)$$

and

$$e^{-\lambda}=1-\frac{8\pi}{3}\rho r^2, \qquad (19)$$

in the usual notation. Thus the condition (3) implies (17) and hence (19). Schwarzschild's internal solution for which ρ is constant satisfies (19) and the corresponding space-time is conformal to flat space-time.

3. THE CONDITION, $C^h_{\ ilj,h}=0$.

Let us now examine the relation (9) for the field given by (10). It is obvious from (6) and (7) that, if (9) is true,

$$R_{il,j}-R_{ij,l}+\tfrac{1}{6}(g_{ij}R_{,l}-g_{il}R_{,j})=0. \qquad (20)$$

For all fields satisfying the equations,

$$R_{ij}=0, \qquad (21)$$

(20) is satisfied and hence (9). For fields occupied by matter,

$$R_{ij}-\tfrac{1}{2}Rg_{ij}=-8\pi T_{ij}, \qquad (22)$$

and (20) now becomes

$$T_{ij,l}-\tfrac{1}{3}g_{ij}T_{,l}=T_{il,j}-\tfrac{1}{3}g_{il}T_{,j}. \qquad (23)$$

Multiplying both the sides of (23) by g^{ij} and summing up, we have

$$T^j_{l,j]}=0. \qquad (24)$$

Thus (23) implies the identities (24) which follow from the field equations (22). But the additional content of (23) exercises a severe restriction on (10) which we now proceed to investigate.

Although there are 64 different equations contained in (23) most o them are found to be identities—the independent equations being only the following four :

$$re^{-\lambda}\mathrm{T}_{11}=\frac{\partial}{\partial r}\mathrm{T}_{22}-\frac{1}{r}\mathrm{T}_{22}+\tfrac{1}{3}r^2\frac{\partial \mathrm{T}}{\partial r}, \quad . \quad . \quad . \quad . \quad . \quad (25)$$

$$re^{-\lambda}\mathrm{T}_{14}=\frac{\partial}{\partial t}\mathrm{T}_{22}+\tfrac{1}{3}r^2\frac{\partial \mathrm{T}}{\partial t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

$$\frac{\partial}{\partial r}\mathrm{T}_{44}-\tfrac{1}{3}e^{\nu}\frac{\partial \mathrm{T}}{\partial r}-\frac{\partial \mathrm{T}_{14}}{\partial t}=\tfrac{1}{2}\dot{\lambda}\mathrm{T}_{14}+\tfrac{1}{2}\nu'\mathrm{T}_{44}-\tfrac{1}{2}e^{\nu-\lambda}\nu'\mathrm{T}_{11}-\tfrac{1}{2}\dot{\nu}\mathrm{T}_{14}, \quad . \quad (27)$$

$$\frac{\partial}{\partial r}\mathrm{T}_{14}-\tfrac{1}{3}e^{\lambda}\frac{\partial \mathrm{T}}{\partial t}-\frac{\partial \mathrm{T}_{11}}{\partial t}=\tfrac{1}{2}\lambda'\mathrm{T}_{14}+\tfrac{1}{2}e^{\lambda-\nu}\dot{\lambda}\mathrm{T}_{44}-\tfrac{1}{2}\dot{\lambda}\mathrm{T}_{11}-\tfrac{1}{2}\nu'\mathrm{T}_{14}. \quad . \quad (28)$$

For a perfect fluid,

$$\mathrm{T}_{\mu\nu}=(p+\rho)\nu_{\mu}\nu_{\nu}-pg_{\mu\nu}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

and

$$g^{\mu\nu}\nu_{\mu}\nu_{\nu}=1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

ν_{μ} being the flow vector. For (10) the field equations give

$$\nu_2=0, \qquad \nu_3=0, \qquad \mathrm{T}_{22}=pr^2, \qquad \mathrm{T}_{33}=pr^2\sin^2\theta \quad . \quad . \quad . \quad (31)$$

and from (29) and (30) we have

$$\mathrm{T}=\rho-3p. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

The equations (25) to (28) now become

$$-(p+\rho)\nu_1\nu^1=\tfrac{1}{3}\rho' r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

$$-(p+\rho)\nu^1\nu_4=\tfrac{1}{3}\dot{\rho} r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

$$e^{\nu}\left[p'+\rho'+\frac{r\rho''}{3}+\tfrac{1}{2}\nu'\left(p+\rho+\frac{2\rho' r}{3}\right)\right]=\tfrac{1}{3}e^{\lambda}[\dot{\rho} r+\tfrac{1}{2}\dot{\rho} r(3\dot{\lambda}-\dot{\nu})], \quad . \quad (35)$$

$$\tfrac{1}{3}\dot{\rho} r(\lambda'+\nu')=\dot{\lambda}(p+\rho+\tfrac{2}{3}\rho' r), \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

respectively. From (33) and (34) it follows that

$$(p+\rho)\nu_4\nu^4=p+\rho+\tfrac{1}{3}\rho' r, \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

and, on eliminating ν_1 and ν_4 with the help of (29), we get

$$\rho'\left[\rho'+\frac{3}{r}(p+\rho)\right]=\dot{\rho}^2 e^{\lambda-\nu}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

The problem now is to determine λ and ν in (10) subject to the five field equations, (13) to (16), and the four equations, (33) to (36). Actually (35) and (36) correspond to $l=1$ and $l=4$ respectively in (24) and are thus implicit in the field equations themselves. In practice it is convenient to use (35) instead of (14). From (37) and (15) we get

$$e^{-\lambda}=1-\frac{8\pi\rho r^2}{3}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

which is the same as (19). Hence for (10) the conditions (3) and (9) are satisfied by the same value of $e^{-\lambda}$. (13), (33) and (36) are satisfied by

$$8\pi(p+\tfrac{1}{3}\rho+\tfrac{1}{3}\rho' r)=\frac{\nu'}{r}\left(1-\frac{8\pi}{3}\rho r^2\right). \quad . \quad . \quad . \quad . \quad (40)$$

The equations (16), (34) and (39) are mutually consistent.

The solution is ultimately provided by (39), which gives λ, and by (35), (38) and (40) which give p, ρ and ν. All these four equations are found to be satisfied by Schwarzschild's internal solution for which ρ is constant. It is interesting to note that p does not appear explicitly in any one of the equations. Either $\rho'>0$ or, as is clear from (38), $\rho'<0$, in which case,

$$\rho'+\frac{3}{r}(p+\rho)\leqslant 0,$$

or

$$\rho'\leqslant -\frac{3}{r}(p+\rho).$$

Stability demands that ρ' be zero or negative. These limitations on density distributions indicate that only Schwarzschild's solution ($\rho'=0$, $\dot{\rho}=0$) is of physical interest under the restriction (9) on (10) when $R_{ij}\neq 0$. If $\rho'<0$ and hence

$$\rho'\leqslant \quad \frac{3}{r}(p+\rho),$$

we have

$$M'r\leqslant -12\pi\int_0^r pr^2\,dr\leqslant 0,$$

where

$$M=4\pi\int_0^r \rho r^2\,dr.$$

This case is of no physical interest since $M'\leqslant 0$.

REFERENCES.

EISENHART, L. P., 1926, *Riemannian Geometry*, 89–92.

TOLMAN, R. C., 1934, *Relativity, Thermodynamics and Cosmology*, 251.

XIV. *The Fluorescence of Silver Halides at Low Temperatures.* Part I. *The Pure Halides* *.

By G. C. Farnell, P. C. Burton and R. Hallama †,
Research Laboratories, Kodak Limited, Wealdstone, Middx.

[Received November 25, 1949.]

[Plates III.–VI.]

Abstract.

Fluorescence emission spectra of AgCl, AgBr and AgI, believed to be characteristic of the pure solids, have been recorded at liquid nitrogen and liquid hydrogen temperatures. The spectra are all in the visible and lie just to the long wavelength side of the absorption bands of the respective halides. No resolution of the spectra was observed down to the lowest temperatures; no pure halide specimen emitted any phosphorescence or thermal afterglow. The two different structures of AgI capable of existence at ordinary and low temperatures (wurtzite and zincblende) gave slightly different emission spectra.

The temperature at which the fluorescence just became visible was about 180° K. for AgCl and AgI, but only 77° K. for AgBr. Excitation of the fluorescence could be effected by any wavelength which the halide was capable of absorbing.

All three halides emitted an extra band in the red after being exposed to light action at ordinary temperature, or slightly decomposed by heat treatment, so as to give a dispersion of free silver in the halide. Large quantities of impurity ($\sim$1–3 M. per cent) were required in order to alter appreciably the characteristic emission of the basic silver halide.

The low-temperature fluorescence of the pure halides is believed to be due to the direct recombination of excited electrons with positive holes with the emission of visible radiation, instead of the emission of thermal quanta as at higher temperatures.

1. Introduction.

This work was prompted by Meidinger's investigation of the low-temperature fluorescence of suspensions of silver chloride, bromide, iodide, and bromoiodide in gelatin, prepared in the manner of photographic emulsions (Meidinger 1939, 1940). It was uncertain whether the pure silver halides would fluoresce in the absence of gelatin: Randall observed that silver chloride shows a bright blue-green fluorescence at 77° K., but thought that some impurity was responsible. We find that all three halides fluoresce strongly in the pure state at a sufficiently low temperature, that is, they show the so-called "fundamental" or "lattice" emission.

---

* Communication No. 1294H from the Kodak Research Laboratories.

† Now at the University of Helsinki.

Until quite recently it was thought that pure solids, apart from manganese, rare-earth, and uranyl compounds having screened incomplete shells of electrons, did not show fluorescence (Seitz 1939; Mott and Gurney 1940). Fluorescence in the pure state, however, has been reported for NaCl, KCl, NaBr, KBr, (Kudriavzewa 1934) ZnS, ZnO and CdS (Kröger 1940), and on theoretical grounds one expects pure substances in general to fluoresce at low enough temperatures: what calls for explanation is not so much the fact of fluorescence as the mechanism by which it is quenched at high temperatures. The main interest of the present work lies in the finding that AgCl, AgBr and AgI have to be added to the above list. All these substances show a low-temperature fluorescence band lying on the long-wavelength flank of the absorption band.

The present paper (Part I.) is concerned with the individual silver halides: Part II. will deal with mixed silver-halide crystals. The results to be described consist mainly of emission spectra, though some observations were made of excitation spectra and the temperature-variation of the intensity of emission. Except where otherwise stated, all results were obtained at liquid-nitrogen temperature (77° K.).

2. Specimen Preparation and Experimental Technique.

(i) *Specimen preparation.*

The specimens were prepared using distilled water and A.R. grade, or purer, reagents, *e.g.* the $AgNO_3$ employed in photographic emulsion manufacture which is purer than A.R. grade. In the precipitations 0·1 M. solutions were employed, as the washing of the precipitate is inefficient with concentrated solutions (Frankenburger, 1923). During the preparation and handling, Kodak Wratten Series "O" safelights were used for darkroom illumination; occasionally preparation was carried out in darkness. No exposure to actinic light was allowed until the specimen was at low temperatures.

Usually equivalent amounts of silver nitrate and metal halide solutions were mixed, *e.g.* 50 ml. 0·1 M. KCl and 50 ml. 0·1 M. $AgNO_3$ (yielding "equivalent crystals"). "Excess Ag" specimens were made by pouring the halide solution into an $AgNO_3$ solution containing 10 M. per cent more than the equivalent quantity of solute, thus ensuring an excess of silver ion throughout the precipitation. The converse procedure was employed to make "excess halide" specimens. All precipitates, particularly the "excess" ones, were thoroughly washed and dried before testing.

To eliminate the possible influence of inclusions of solution in the precipitate certain samples were shaken in water for some hours. In the cases of AgCl and AgBr this yielded large grains of about 1 mm. in diameter, not single crystals but hard compact aggregates, which, according to Frankenburger, have no inclusions. Further special preparations of various specimens will be described later, where appropriate.

(ii) *Experimental technique.*

The specimens were usually loaded into the lower closed end of a length of $\frac{3}{8}$-inch glass tubing, which was immersed in a transparent Pyrex or quartz Dewar flask containing liquid nitrogen. They were excited by the radiation of a 125 w. mercury high-pressure arc transmitted by a Wratten 18A filter (3130 Å.–4047 Å.). Excitation by the Hg lines transmitted by a 10 cm. thickness of a compound solution of 41·5 g. $CoSO_4.7H_2O$ and 145 g. $NiSO_4.6\text{–}7H_2O$ in 1 litre of distilled water (2480 Å.–3341 Å.) (Bowen, 1935) was found to give the same results. In this case specimens were mounted as layers on the lower end of a glass strip, a thin layer of gelatin being employed as adhesive. Specimens, immersed in the quartz Dewar flask of liquid nitrogen, could then be irradiated without the exciting radiation having to pass through glass.

The fluorescent light was focused on to the slit of a Hilger Medium quartz spectrograph (with occasional recourse to the Hilger Constant Deviation glass instrument), and the spectrum recorded on Kodak P.1200 or P.1500 plates. A Wratten No. 2 or 2A filter was mounted at the condensing lens to eliminate the ultra-violet radiation scattered in the direction of the spectrograph. The glass condensing lens eliminated the shorter wavelength ultra-violet when the Bowen filter was employed. Direct Hg spectra were always recorded with the fluorescence spectrograms, for calibration purposes. With the brightest specimens and a slit width of $\frac{1}{10}$ mm. an adequate exposure time was 1–2 minutes. Precisely the same technique was employed for recording the emission of specimens at liquid hydrogen temperature (20° K.).

Certain spectra were examined with the aid of a photomultiplier cell. The emission of the specimen was allowed to pass right through the quartz spectrograph, at the farther side of which it was further dispersed by a large replica grating and brought into focus; the linear extension of the spectrum was increased by a factor of three. The photomultiplier cell with a narrow slit in front of its cathode, was moved along a graduated scale in the plane of the spectrum, and current readings were taken at various regular intervals. With the help of line-spectrum calibration of the scale, and the spectral-sensitivity curve of the photomultiplier cell, it was possible to obtain a curve of relative intensity against wavelength for the fluorescent light. This was much more convenient than correcting for spectral sensitivity of the panchromatic plates employed in the photographic recordings.

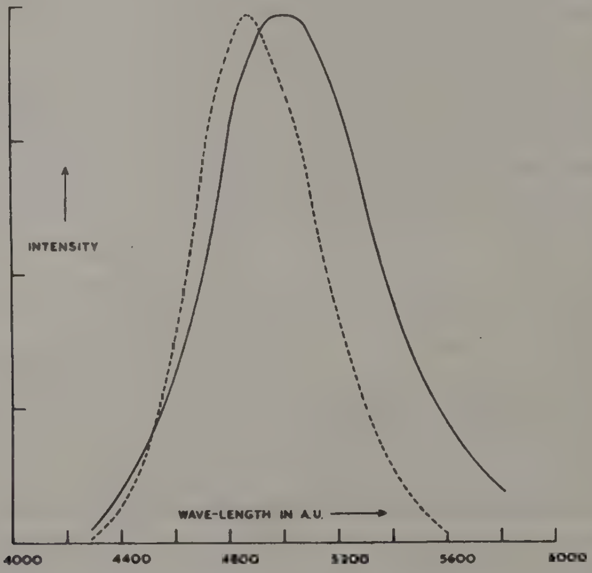
The excitation of various Hg lines and groups were observed by dispersing radiation from a high-intensity Hg source with the dispersing system employed with the photomultiplier. A small quartz Dewar flask containing the specimen in a tube immersed in liquid nitrogen was placed in the plane of the Hg spectrum and the effects of separate lines observed through a grating, which dispersed at right angles to the direction of dispersion of the exciting system.

3. SILVER CHLORIDE.

(i) *Emission spectra.*

All precipitated specimens at 77° K. emitted a broad structureless band extending from 4350 Å. to 5800 Å. (Pl. III., fig. 1, *a*). The colour of the fluorescence was blue-green. Measurements with the photomultiplier arrangement showed that it was a single smooth band with a peak at 5000 Å. (fig. 1). The extent and position of the band agrees with Randall's results (Randall 1939). That it cannot be ascribed to any impurity in the precipitating reagents seems clear from the fact that specimens precipitated from eleven different metal halides (*e. g.* NH_4, Fe, Co, Ni, Mn, Ca) and several different specimens of $AgNO_3$ and Ag_2SO_4 gave identical spectra. Some examples are shown in Pl. III., fig. 1, *b*, to *e*.

Fig. 1.



The spectral distribution of fluorescence of AgCl at low temperatures. Intensity in arbitrary energy units :

Solid curve—specimen at 77° K.
Dashed curve—specimen at 145° K.

Ordinates of latter curve adjusted to give same intensity at maximum, so as to show more clearly the shift of the band to shorter wavelengths on increase of the specimen temperature.

To eliminate the possibility that structural defects in the precipitate were responsible for the low temperature emission, single crystals of AgCl (1–2 mm. in size) were grown from ammonia solution and examined. Their low-temperature fluorescence was just as intense as that of the precipitates. It was still possible that defects associated with an excess of silver were leading to the observed emission, but this was discounted by the fact that annealing the large crystals at 400° C. for five hours in an atmosphere of chlorine, and then cooling slowly to room temperature (which would also tend to eliminate structural defects) gave specimens which were even more strongly fluorescent than before the annealing treatment. A specimen prepared by the action of pure, dry chlorine on Hilger "Specpure" silver gave an emission spectrum identical with those already recorded.

If a specimen was melted in the open air, either in a platinum or porcelain crucible, so that, in the reducing atmosphere of the Bunsen flame, slight decomposition took place, with the formation of free silver, the emission was always found to include a band in the red (5800 Å. to 6450 Å. Pl. III., fig. 1, *f* and *g*). This red band could also be induced by a short exposure of the specimen to ultra-violet radiation at ordinary temperatures, which led to the formation of "print-out" silver. Careful tests showed that the red emission, which was observed with all halide specimens after appropriate preparation or treatment, could only be excited by radiation which was absorbed by the halide, *i. e.* there was no specific excitation band of the silver impurity. It was induced more readily in an "excess silver" specimen. The red emission is obviously associated with the silver excess, but whether it is due to small specks or atoms of silver acting as impurity centres, or to the formation of F-centres, is at present an open question. Melts showing the red fluorescence also gave a weak red phosphorescence and a thermal afterglow occurring in two bursts, the first having a green and the second a red colour.

(ii) *Impurity activation.*

The results of these experiments are important in considering the validity of the claim that pure AgCl is fluorescent at low temperatures. Mn, Mg, Cu and Pb (as chlorides) were incorporated in the AgCl as impurities (by melting in halogen atmosphere), in quantities ranging from 10^{-4} M. per cent to 5 M. per cent. It was found that comparatively large quantities of impurity ($\sim$1–3 M. per cent) were required in order to alter the normal AgCl emission appreciably. Such changes as occurred usually consisted of an extension of the emission towards the red; the emission spectogram of a specimen containing 5 M. per cent Mg is shown as an example (Pl. III., fig. 1, *h*). Specimens containing $\sim$1 M. per cent Mn showed a fairly strong red emission band to be ascribed to the Mn impurity (Pl. III., fig. 1, *i*); at 5 M. per cent Mn the normal AgCl emission had practically been suppressed and replaced by a strong red Mn band (5920 Å.–6450 Å.; Pl. III., fig. 1, *j*).

The specimens which contained $\sim$1 M. per cent Mn were particularly interesting in that their colour of emission changed from green to red as the intensity of excitation was reduced, and showed a change in the same direction as their temperature was raised. This phenomen is analogous to that observed in the case of ZnS+Cu, or ZnS+Mn (Gisolf and Kröger, 1939), and can be explained by the mechanism put forward by Schön (1942) and Klasens (1946). It depends essentially on the fact that the energy gap between the ground state of the AgCl centres and the $3p$ band of AgCl is less than the corresponding gap for the Mn centres. An analysis based on this and certain other simplifying assumptions (for which reference must be made to the original sources) leads to expressions which are in qualitative agreement with the phenomena observed. Perhaps the most interesting fact to be deduced from the above observations is that positive holes are mobile in AgCl at low temperatures, at least over a distance of a few lattice spacings.

(iii) *The influence of temperature.*

Measurements with a copper-constantan thermocouple have shown that the fluorescence of AgCl commences at a temperature of 180° K. From this temperature down to 20° K. the emission band (or rather the peak) moves over towards longer wavelengths. At 150° K. the emission appears blue in colour, at 77° K. blue-green, and green at 20° K. This is well illustrated by the three spectrograms shown in Pl. III., fig. 2, *a–c*, recorded at the three specimen temperatures noted above. (The emission at 20° K. was recorded with the glass spectrograph and thus the spectrogram is on a different scale from the other two.) The shift to shorter wavelengths with increasing temperature is also illustrated in fig. 1.

(iv) *Excitation.*

Tests with the arrangement described in § 2, (*ii*), have shown that AgCl is strongly excited by all Hg lines in the ultra-violet region, but only weakly by the Hg 4047 Å. line. The absorption edge of AgCl is at about 4300 Å. and it is thus apparent that the fluorescence of AgCl is excited by all wavelengths which it absorbs. The emission excited by wavelengths in the range 2480 Å.–3130 Å. was identical in spectral distribution with that excited by the longer ultra-violet wavelengths.

4. SILVER BROMIDE.

(i) *Emission spectra.*

Ordinary precipitates and open-atmosphere melts of AgBr at 77° K. showed a faint emission band in the red extending from 5700 Å. to 6450 Å. (Pl. IV., fig. 3, *a* and *b*). This red band was strongest in the slightly decomposed melts and the "excess silver" precipitate. Spectroscopic tests with Kodak Infra-Red Extra-Rapid plates showed that it did not appreciably extend beyond 6450 Å. The red emission was suspected of being due to a silver impurity in the various specimens, and to test this, a specimen free from silver was prepared by melting some AgBr in an atmosphere of bromine and then quenching it in distilled water. It was found to be completely non-fluorescent at 77° K. When it had been allowed to "print out" at ordinary temperatures for a few seconds, however, the red emission was again observable at low temperatures. This is fairly conclusive evidence that the red emission is due to the presence of silver.

The "excess bromide" specimen at 77° K. gave an additional faint emission band in the green (Pl. IV., fig. 3, *c*) and it was believed that this represented the emission of pure AgBr, in analogy to the blue-green emission of AgCl, but that a much lower temperature was required in order to observe it at high intensity. This belief was found to be justified when observations were made at a specimen temperature of 20° K. AgBr gave a very strong green emission in a single band extending from 4720 Å. to 5375 Å. with a peak at 5050 Å. This is shown in Pl. IV., fig. 3*d*, along with the red "silver" band which appeared with quite high intensity at 20° K. We believe the green emission band to be characteristic of pure AgBr, although it must be admitted that there is not such an overwhelming weight of evidence in favour of this claim as in the case of AgCl.

(ii) *Impurity activation.*

The most striking result in this section, but one which properly belongs to Part II., was the effect of precipitating small quantities ($\sim$5 M. per cent) of AgI in AgBr so as to form mixed crystals. Such specimens showed a very strong green emission, the spectrum of which was a single band extending from 5150 Å. to 6000 Å. with a peak at 5500 Å. (Pl. IV., fig. 4, *a*). The position agrees with that recorded by Meidinger for the emission of silver bromo-iodide emulsion layers. We believe, for reasons to be given later, that the emission of such specimens is seated in the iodide ion, and that the green band is not a "sensitized" bromide band.

Evidence supporting the latter contention is provided by the fact that inclusions of other mono- and divalent halides, *e. g.*, CuCl, $CuCl_2$, $PbCl_2$, $PbBr_2$, $MgCl_2$, to the extent of 5 M. per cent, had only very weak activating effects. In general they gave specimens showing weak emission bands in the red and yellow. (The red bands were not due to a silver impurity, as the inclusion of impurity by melting was carried out in a bromine atmosphere.) Some examples of the spectra are shown in Pl. IV., fig. 4, *b–d*.

5. Silver Iodide.

(i) *Emission spectra.*

The essential features of an AgI emission spectrum are well illustrated by P.   ., fig. 5, *a*, the specimen in this case being prepared by the action of iodine vapour on spectroscopically pure silver. The main band extends from 4400 Å. to 4600 Å.; the narrow short wavelength band was of variable position, but in the case illustrated is situated at 4220 Å. A red band, in a similar position to the "silver" bands of AgCl and AgBr.

could be induced by exposure of the AgI at ordinary temperatures. Almost all precipitates emitted this red band; it was particularly strong in "excess silver", and aged, precipitates (Pl. V., fig. 5, *b* and *c*). Some specimens also emitted a faint band in the green (Pl. V., fig. 5, *c* and *d*), which was assigned to some impurity, since it was not present in the emission of the purest specimen prepared by the action of iodine on silver, or in the emission of the majority of precipitates prepared with alternative reagents.

A puzzling feature of the first recordings of emission spectra of various AgI specimens was the manner in which the positions of the main band and narrow band altered, depending on the conditions of precipitation, *e. g.*, the narrow band was emitted at shorter wavelengths in the case of "excess iodide" specimens (Pl. V., fig. 5, *e*). It was later realized that these variations were almost certainly due to the fact that at ordinary temperatures AgI is precipitated as a mixture of wurtzite-type (hexagonal) and zincblende-type (cubic) structures (Wilsey 1923), and that different specimens contained different proportions of the two structures. X-ray diffraction patterns confirmed this surmise, and so extensive efforts were made to prepare AgI specimens of one single structure. The methods employed for obtaining the wurtzite structure were :—

(*a*) Precipitation of AgI dissolved in concentrated HI solution by slow addition of ethanol (Helmholz 1935).

(*b*) Rapid precipitation of AgI by dilution of a concentrated solution of AgI in KI (Bloch and Möller 1931).

(*c*) Slow cooling of an AgI melt (Bloch and Möller 1931).

(*d*) Slow crystallization of AgI from a dry piperidine solution by absorption of atmospheric water vapour by the latter. This method yielded soft aggregates of external hexagonal form, but which were opaque and easily crumbled to a mass of small crystals.

For obtaining the zincblende structure, the methods used were :—

(*e*) Rapid precipitation of AgI by dilution of a concentrated solution of AgI in $AgNO_3$ (Bloch and Möller 1931, Kolkmeijer and Hengel 1934).

(*f*) Pulverization of any hexagonal specimen, *e. g.* pieces of slow cooled melt of (*c*) or the soft aggregates of (*d*), in an agate mortar and pestle.

Debye-Scherrer powder diagrams showed that in no case was a specimen obtained whose structure was completely one or the other type, but some approached 90 per cent structural purity. From these it was possible to distinguish the features characteristic of the emissions of the two different structural types. They are particularly well illustrated by Pl. ., fig. 6, *a* and *b* which show the emission of a piece of slow cooled melt and that of a predominantly cubic precipitate prepared by method (*e*). The narrow band is at a shorter wavelength in the case of the hexagonal specimen, while the main band is at longer wavelengths. The latter fact is even better illustrated by the two low-exposure spectrograms shown in

Pl. V., fig. 6, *c* and *d* of the emission of predominantly cubic and hexagonal precipitates. A further comparison of predominantly cubic and hexagonal precipitates is made in Pl. V., fig. 6, *e* and *f*. It should be noted how intense the red "silver" bands are in the emission of the cubic precipitates; their preparation in all cases involved a heavy excess of silver.

The change over in emission characteristics when a specimen is converted from hexagonal to cubic by method (*f*) is illustrated by Pl. V., fig. 7. The first spectrogram (Pl. V., fig. 7, *a*) is that of the fluorescence of predominantly hexagonal crystals prepared by method (*d*), the second and third (Pl. V., fig. 7, *b* and *c*) those of the same crystals after being ground for 10 and 20 minutes respectively. The growth of the narrow band at longer wavelengths and the shift of the main band towards shorter wavelengths are easily visible. X-ray powder diagrams confirmed a general change-over from hexagonal to cubic structure.

A careful study of all spectrograms and the X-ray powder diagrams of the corresponding specimens showed that all the variations in spectral emission exhibited were due to the various proportions of the two different structural types in the various specimens. For instance, the specimen precipitated in excess iodide solution which gave a narrow band at shorter wavelengths (Pl. V., fig. 5, *e*) than did other precipitates, was found to be predominantly hexagonal in structure.

A puzzling feature of the low-temperature emission of AgI is the fact that the narrow short wavelength emission band falls within the (low-temperature) absorption band of the halide. This is illustrated by Pl. VI., fig. 8, *a*–*c* which contrasts the emissions of predominantly cubic and hexagonal specimens (fig. 8, *a* and *c*) with the absorption spectrum of a mixed specimen (fig. 8, *b*) (tests showed that the wavelength of the absorption edge was sensibly independent of the structure of the AgI specimen). We believe that the main emission band of AgI, and the emission bands of AgCl and AgBr, are due to the direct recombination of the excited electron with a positive hole at low temperatures, this process having a relatively low probability at higher temperatures compared with the probability of radiationless recombination. At the present moment, however, we cannot suggest any mechanism which explains the emission of the narrow AgI band, or why it is emitted in a spectral region where the absorption coefficient is high ($\sim 10^5$ cm.$^{-1}$) (Schell 1911).

Its intensity, however, does appear to be controlled by the absorption of the halide as it was found that the narrow band was much more intense from a specimen which had been sedimented into a compact layer, than from the same specimen mounted in a tube as a loose, fine powder. This is illustrated by Pl. VI., fig. 9, *a* and *b*. The reason why the self absorption should be less in the first case is not at all obvious, however. The intensity of the narrow band was particularly high in the case of emissions from smooth surfaces of melts, and was much reduced when such specimens were broken down with little change in structure. It thus appears that the narrow band emission cannot be assigned to surface ions.

(ii) *The influence of temperature.*

The fluorescence of hexagonal specimens became visible at about 195° K. and that of cubic specimens at a slightly lower temperature. At 77° K. hexagonal specimens showed an appreciably higher fluorescence efficiency. Pl. VI., fig. 10 compares the emissions of a mixed AgI specimen at 77° K. (*a*) and 20° K. (*b*), identical exposure conditions being employed in each recording. It will be seen that the fluorescence is much more intense at the lower temperature, and that the spectrum has shifted and extended towards shorter wavelengths. The extension towards shorter wavelengths is probably connected with the retraction of the absorption edge on reduction of the temperature. The narrow bands do not appear to be shifted, and have undergone little change in intensity.

(iii) *Excitation.*

All Hg lines up to 4047 Å. were found to excite specimens of AgI strongly, but not lines beyond this wavelength. The absorption edge of AgI at low temperatures is at 4250 Å. and thus, as in the case of AgCl, we see that fluorescence is stimulated by all wavelengths which are capable of being absorbed. This result is indeed to be expected if the fluorescence observed is characteristic of the pure solid.

6. PHOTOGRAPHIC EMULSIONS.

Some attention was paid to the fluorescence of photographic emulsions at low temperatures. The emission spectra corresponded to a simple superposition of the emissions of their silver-halide contents and gelatin. The latter, investigated separately, was found to show a blue fluorescence at both ordinary and low temperatures ; the spectrum extended over the whole of the visible region. It also showed a moderately strong phosphorescence at low temperatures, which was reproduced in the emulsions.

No differences in emission were observed between dye-sensitized and unsensitized emulsions of otherwise identical composition. Photographic desensitizers were also found to have no appreciable influence on the fluorescence. It also appeared that emulsion impurities, such as silver sulphide specks on the grain surfaces, and silver-gelatin complexes, etc., had no important activating or de-activating effects.

Meidinger has reported that the efficiency of fluorescence of bromo-iodide emulsions (96 per cent AgBr, 4 per cent AgI) depends strongly on grain size at a given temperature. At 77° K. a coarse-grain emulsion gave a strong fluorescence having a quantum efficiency of about 0·9, a series of finer-grain emulsions gave progressively lower efficiences, and a Lippmann-type emulsion of very fine grain showed no fluorescence at all. We have been unable to confirm this result : we have examined a number of bromo-iodide emulsions of various grain sizes, including a specially prepared series having the same constitution and a wide range of grain sizes, and several transparent Lippmann-type emulsions, and all showed about the same high intensity of fluorescence. A speculative explanation of Meidinger's results on fine-grain emulsions would be that he employed a method of precipitation that enabled the relatively insoluble AgI to precipitate before the AgBr with little formation of genuine mixed crystals.

The fact that we recorded identical spectra for the emissions of bromo-iodide precipitates and bromo-iodide emulsions (except for the weak overlying gelatin emission spectrum in the latter case) shows that the green fluorescence of the latter is not due to activation of the bromo-iodide by silver sulphide specks, as has been suggested by Vassy and Vassy (1948).

7. CONCLUSIONS.

The experimental results which have been described show that the silver halides are fluorescent in the pure state at low temperatures. Besides the evidence provided by the many alternative preparations of the halides, the position of the emission band of each halide on the flank of its absorption band suggests that the fluorescence is of the "fundamental" or "lattice" type. We can explain our results in terms of the following simple model. If the system in the excited state (reached directly by absorption, or after the electron has spent some time in the conduction band) receives sufficient thermal energy to raise it from the minimum of the state to a point where it can cross over to a high vibrational level of the ground state, it will accomplish the return to the bottom of the ground state with the emission of vibrational quanta. At low temperatures however, where the probability per unit time of the above process has fallen to a low or negligible value, the excited electron will return directly to the ground state (recombination with a positive hole) with the emission of radiation. It will be shown in Part II. how the results for the variation of fluorescence efficiency of silver bromo-iodide fit in with the simple explanation given above.

We expect that careful investigation will show that many pure crystals exhibit a low-temperature fluorescence analogous to that of the silver halides. We have observed a fluorescence of the cuprous halides which is probably of this type. Our interest was centred on the silver halides in the hope that information would be obtained of some use in photographic latent-image theory. It is known that a developable latent image can be formed in a photographic emulsion by exposure at or below liquid nitrogen temperature; it is supposed that the photoelectrons released during the exposure are trapped mainly in shallow traps, and released on warming up, when also the ionic movements necessary for latent-image formation can take place (Berg 1939). It has been found that in all cases there is a fall in sensitivity at low temperatures as compared with that at ordinary temperatures. Some of this fall is no doubt due to the wastage of the energy of the exposure as fluorescence.

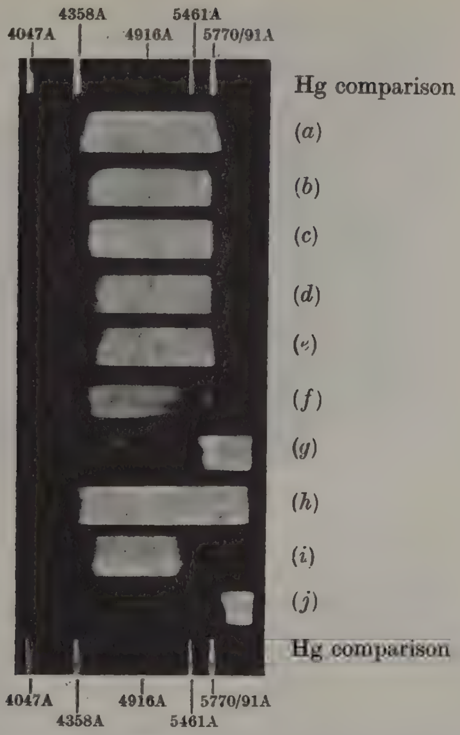
ACKNOWLEDGMENTS.

We are indebted to Professor F. E. Simon, C.B.E., F.R.S., of the Clarendon Laboratory, Oxford, for his assistance in enabling us to carry out the observations at liquid hydrogen temperature. Our thanks are due also to Mr. A. H. Fox of these laboratories who recorded many of the X-ray powder diagrams of our specimens, and Mr. D. J. Yates who rendered assistance in some of the experiments.

REFERENCES.

BERG, W. F., 1939, *Trans. Faraday Soc.*, **35,** 445.
BLOCH, R., and MÖLLER, H., 1931, *Z. phys. Chem.*, **152,** 245.
BOWEN, E. J., 1935, *J. Chem. Soc.*, 76.
FRANKENBURGER, W., 1923, *Z. phys. Chem.*, **105,** 273.
GISOLF, J. H., and KRÖGER, F. A., 1939, *Physica*, **6,** 1101.
HELMHOLZ, L., 1935, *J. Chem. Phys.*, **3,** 740.
KLASENS, H. A., 1946, *Nature*, **158,** 483.
KOLKMEIJER, N. H., and HENGEL, J. W. A., 1934, *Z. Kristallogr.*, **88,** 317.
KRÖGER, F. A., 1940, *Physica*, **7,** 1.
KUDRIAVZEWA, W., 1934, *Z. Phys.*, **90,** 489.
MEIDINGER, W.,1939, *Phys. Z.*, **40,** 517.
1940, *Phys. Z.*, **41,** 277.
MOTT, N. F., and GURNEY, R. W., 1940, p. 204, *Electronic Processes in Ionic Crystals.* (Oxford : The University Press).
RANDALL, J. T., 1939, *Trans. Faraday Soc.*, **35** (1), 2.
SCHELL, C., 1911, *Ann. Phys.*, **35,** 695.
SCHÖN, M., 1942, *Z. Phys.*, **119,** 463.
SEITZ, F., 1939, *Trans. Faraday Soc.*, **35** (1), 74.
VASSY, A., and VASSY, E., 1948, *C.R. Acad. Sci., Paris*, **226,** 1183.
WILSEY, R. B., 1923, *Phil. Mag.*, **46,** 487.

FIG. 1.

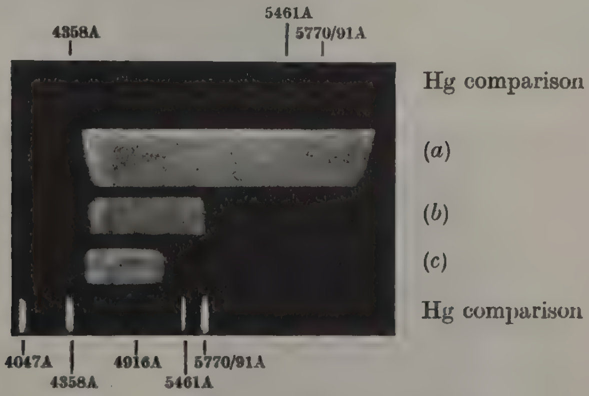

4358A
5461A
4047A
4916A
5770/91A
Hg comparison
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
Hg comparison
4047A
4916A
5770/91A
4358A
5461A

Emission spectra of silver chloride specimens.

(a) AgCl from $KCl+AgNO_3$.
(b) ,, ,, $CaCl_2+AgNO_3$.
(c) ,, ,, $MgCl_2+AgNO_3$.
(d) ,, ,, $CoCl_2+AgNO_3$.
(e) ,, ,, $MnCl_2+AgNO_3$.

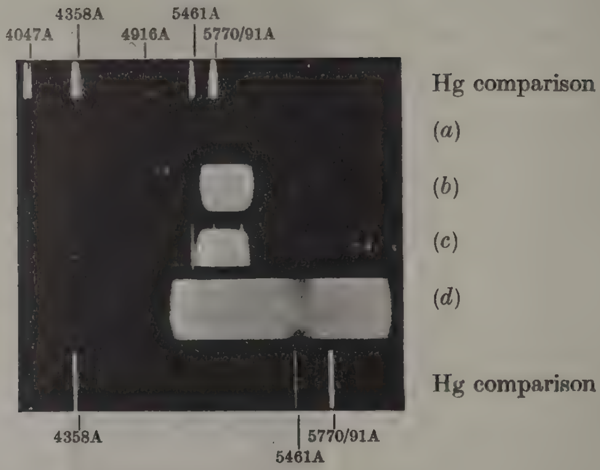
(f) AgCl *slightly decomposed* (*free* Ag *present*).
(g) AgCl *heavily decomposed* (,, ,, ,,).
(h) $AgCl+5$ *mol. per cent* $MgCl_2$.
(i) $AgCl+1$ *mol. per cent* $MnCl_2$
(j) $AgCl+5$ *mol. per cent* $MnCl_2$.

FIG. 2.


5461A
4358A
5770/91A
Hg comparison
(a)
(b)
(c)
Hg comparison
4047A
4916A
5770/91A
4358A
5461A

Emission spectra of silver chloride at various temperatures.
(a) 20° K. (*glass spectrograph*). (b) 77° K. (c) 150° K.

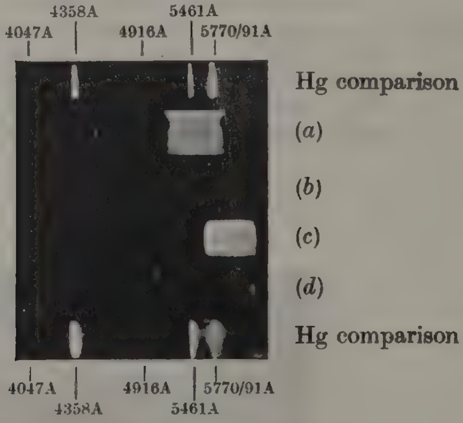
Fig. 3.


4358A
5461A
4047A
4916A
5770/91A
Hg comparison
(a)
(b)
(c)
(d)
Hg comparison
4358A
5770/91A
5461A

Emission spectra of silver bromide specimens.

(a) *" Normal " precipitate* (77° K.).
(b) *Slightly decomposed melt* (77 °K.).
(c) *" Excess bromide " precipitate* (77° K.).
(d) *" Normal " precipitate* (20° K. ; *glass spectrograph*).

Fig. 4.


4358A
5461A
4047A
4916A
5770/91A
Hg comparison
(a)
(b)
(c)
(d)
Hg comparison
4047A
4916A
5770/91A
4358A
5461A

Impurity activated silver bromide.

(a) AgBr+5 *mol. per cent* AgI.
(b) AgBr+5 *mol. per cent* $PbCl_2$
(c) AgBr+5 *mol. per cent* $MgCl_2$.
(d) AgBr+2 *mol. per cent* $MnCl_2$.

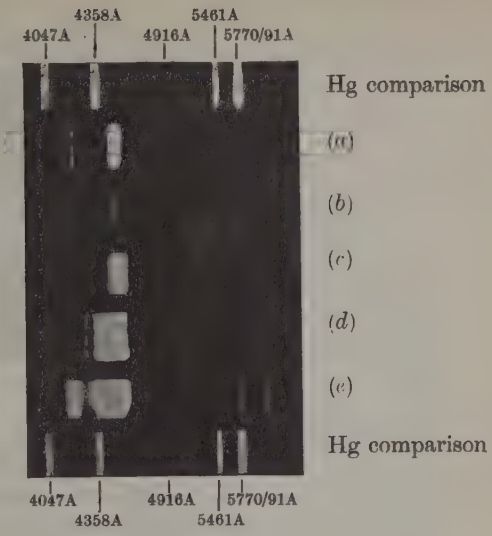

4047A
4358A
4916A
5461A
5770/91A
Hg comparison
(a)
(b)
(c)
(d)
(e)
Hg comparison
4047A
4358A
4916A
5461A
5770/91A

Fig. 5.

Emission spectra of AgI specimens.

- (a) *From action of* I_2 *on* Ag.
- (b) *" Excess silver " specimen.*
- (c) *" Excess silver " specimen, emitting band in green.*
- (d) *From* $HI+AgNO_3$*, emitting band in green.*
- (e) *" Excess iodide " specimen.*

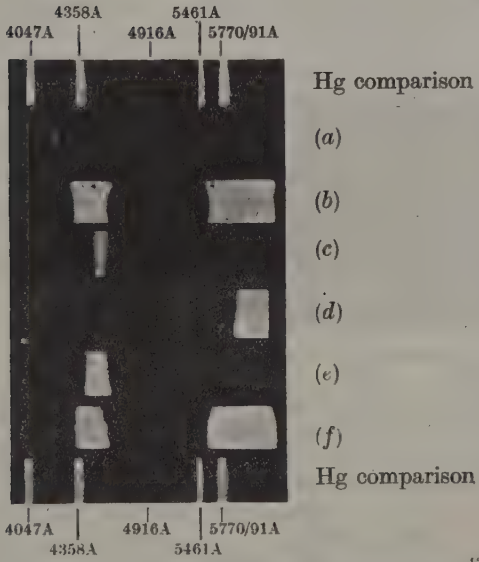

4047A
4358A
4916A
5461A
5770/91A
Hg comparison
(a)
(b)
(c)
(d)
(e)
(f)
Hg comparison
4047A
4358A
4916A
5461A
5770/91A

Fig. 6.

Comparison of the emission spectra of wurtzite (hexagonal) and zincblende (cubic) specimens of AgI.

- (a) *Hexagonal* AgI ; *slow cooled melt.*
- (b) *Cubic* AgI ; *rapid dilution of concentrated solution of* AgI *in* $AgNO_3$.
- (c) *Hexagonal precipitate of* AgI. } *Low exposures.*
- (d) *Cubic precipitate of* AgI. } *Low exposures.*
- (e) *Hexagonal precipitate.* } *Different preparations to those noted above.*
- (f) *Cubic precipitate.* } *Different preparations to those noted above.*

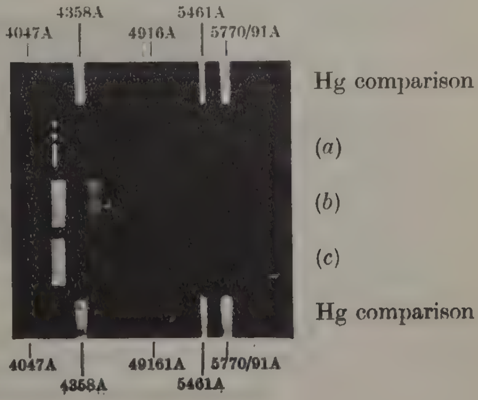
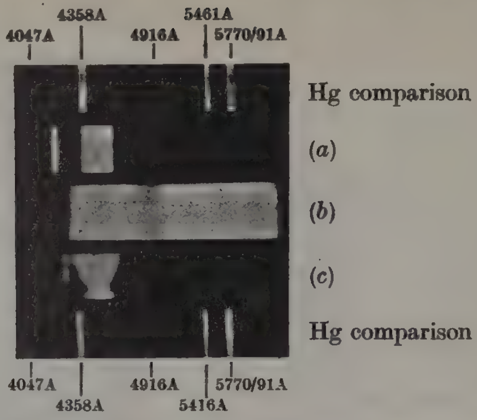

4047A
4358A
4916A
5461A
5770/91A
Hg comparison
(a)
(b)
(c)
Hg comparison
4047A
4358A
4916A
5461A
5770/91A

Fig. 7.

The change in emission spectrum of AgI on change of structure.

- (a) *Hexagonal* AgI *prepared from piperidine solution.*
- (b) *Same specimen after partial conversion to cubic structure by grinding.*
- (c) *Further conversion to cubic structure.*

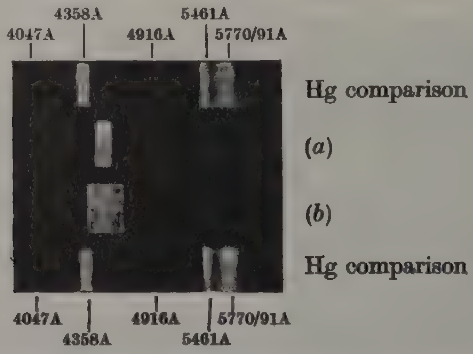
Fig. 8.


4358A
5461A
4047A
4916A
5770/91A
Hg comparison
(a)
(b)
(c)
Hg comparison
4047A
4916A
5770/91A
4358A
5416A

Comparison of the emission spectra of cubic and hexagonal specimens with the absorption spectrum of AgI at 77° K

(a) *Predominantly hexagonal precipitate.*
(b) *Absorption spectrum of mixed precipitate.*
(c) *Predominantly cubic precipitate.*

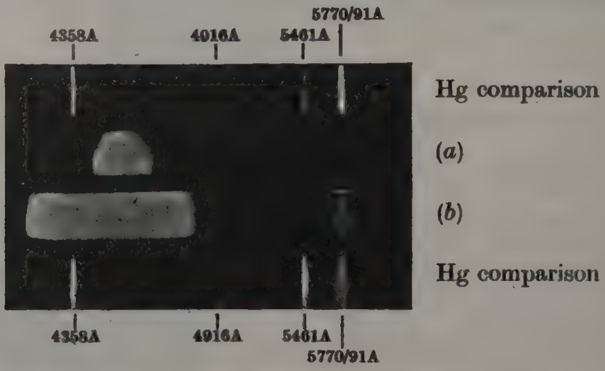
Fig. 9.


4358A
5461A
4047A
4916A
5770/91A
Hg comparison
(a)
(b)
Hg comparison
4047A
4916A
5770/91A
4358A
5461A

Comparison of emission spectra from a compact layer and a loose powder of a particular AgI specimen.

(a) *Compact layer.*
(b) *Same specimen in loose powder form (longer exposure employed in recording).*

Fig. 10.


5770/91A
4358A
4916A
5461A
Hg comparison
(a)
(b)
Hg comparison
4358A
4916A
5461A
5770/91A

Emission spectra of an AgI specimen at 77° K. and 20° K.

(a) 77° K. (b) 20° K. (*same exposure conditions as in* (a)).

Glass spectrograph employed for recording the spectra.

XV. *Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.*—Part III. *Nature of the Shower Particles.*

By P. H. Fowler,
The H. H. Wills Physical Laboratory, University of Bristol *.

[Received December 21, 1949.]

Summary.

A new method of measuring the scattering of charged particles in their passage through photographic emulsions has been developed. It has been employed to determine the mass of charged particles emerging from nuclear "explosions" produced by cosmic radiation. Of the mesons among the "shower" particles, produced in these explosions with momentum less than about 180 MeV./c, it is shown that at least 90 per cent are π-particles; and that less than 5 per cent, and possibly none, are μ-particles, electrons or other particles of small mass. If mesons of greater mass are sometimes created in nuclear interactions with an energy less than 300 MeV., they must be produced with a frequency less than 1 per cent of the protons and less than 5 per cent of the π-particles. The observations therefore prove that the great majority of the charged particles created in nuclear interactions are π-particles.

The greatly increased rate of measurement with the new method and the fact that it allows the determination of the scattering of particles of much greater energy than has been possible hitherto will enable a detailed study of the nuclear transmutations produced by protons and α-particles with energies in the interval from 10^9 to 10^{10} eV. to be made.

Introduction.

In Part II. of the present series of papers (Camerini *et al.*; 1949), which will be referred to as (II.), it was shown that, in favourable cases, it is possible to determine the rest-mass and the energy of fast-charged particles, emitted during the explosive disintegration of nuclei, by observations of the grain-density in the tracks and the deviations in them due to multiple Coulomb scattering. The essential principles of the methods employed for measuring the scattering have been described by Goldschmidt-Clermont *et al.* (1948) and by Davies *et al.* (1949).

---

* Communicated by Professor C. F. Powell, F.R.S.†

† It had been intended that Part III. of the present series of papers should contain a discussion of the experimental material given in Parts I. and II. The rapid development of the methods discussed in Part II.—and, in particular, the improvements in the scattering method described by Mr. Fowler in the present paper—allowed us, however, to obtain a more detailed insight into many features of the nuclear disintegrations produced by protons and α-particles of great energy. It therefore seemed desirable to make a more extensive series of investigations, and the final discussion will be postponed to a later paper. (C. F. P.).

It was pointed out in (II.) that the new method is similar in principle to that employed in experiments with Wilson chambers operated in magnetic fields. The average deviation in a track per unit length, $\bar{\alpha}$, gives a measure of the quantity $p\beta/Z$ for the corresponding particle, where p is its momentum, Ze its charge, and β the ratio of the velocity of the particle, v, to that of light, c. On the other hand, if the charge of the particle is known, the grain-density in its track gives a measure of its velocity—and thence the corresponding value of β. The two observations on the same track therefore allow both the mass and the velocity of the particle to be determined if its energy lies in a particular range of values; *viz.*, that its kinetic energy is less than 0·6 $mc.^2$, where m is the rest-mass, and greater than $\sim$15 MeV. The method has the advantage that it is possible to measure the specific ionization of a particle of the cosmic radiation, by observations on the "grain density" in its track in a photographic emulsion, with a much greater precision than by "drop-counts" from photographs taken with "counter-controlled" Wilson chambers. Further, in favourable circumstances, it appears to be possible to determine the quantity $p\beta/Z$, by scattering measurements, with an accuracy comparable with that commonly achieved in experiments on the magnetic deflection of charged particles of the cosmic radiation.

The importance of the new method lies in the fact that it greatly extends the field of application of the photographic method. Hitherto, using this method of detection, determinations of mass have been confined to particles of low energy which reach the end of their range in the emulsion. The new method removes this restriction and, within certain limits, allows us to determine the mass and energy of particles, irrespective of whether or not they escape from the emulsion.

Whilst the early measurements, described in (II.), showed the potentialities of the method, the determination of the scattering parameter, $\bar{\alpha}$, was arduous, for it involved plotting the individual grains along a track from an image thrown on a screen by means of a projection microscope. This feature placed a limit on the field of application of the method; for with the apparatus at our disposal, an extensive analysis, made desirable if observations of great statistical weight were to be obtained, would have occupied some years. It therefore became desirable to develop simplified methods; and we have found that the following, which employs microscopes of conventional design, has the advantage of greatly increasing the rate of measurement, without an appreciable loss in precision, in addition to extending the range of values of the energy of particles which can be determined.

DETERMINATION OF THE SCATTERING PARAMETER.

In order to determine the scattering parameter, $\bar{\alpha}$, for a given track, the plate is placed on the mechanical stage of the microscope, so that the track is approximately parallel to one of the lines of motion of the stage, say the X-axis. The point, y_1, at which the image of the track intersects a hair-line in one of the eye-pieces, is determined by means of an eye-piece scale. The X-motion of the stage is then moved through a standard distance, say 100 μ, and a reading taken of the new position, y_2, of the image of the track relative to the eye-piece graticule. The operations are repeated along the entire length of the track.

The first differences, $S_1=y_1-y_2$, $S_2=y_2-y_3$ etc. give a measure of the inclination i_1, i_2, etc. of the successive chords in the track—and the second differences, S_1-S_2, S_2-S_3, etc. the deviations, D_1, D_2, etc. between successive chords. The mean value of $|D|$, $\frac{\Sigma_n|D|}{n}=\overline{D}$, taken without regard to sign, gives a measure of the scattering parameter, $\bar{\alpha}$, from which the quantity $p\beta/Z$ can be evaluated.

For the successful application of the method, it is necessary that any apparent deviations in the track due to mechanical imperfections of the stage motions shall be compared with the deviations due to scattering. It has been found that satisfactory performance in this respect is given by Type M4000 microscopes by Cooke, Troughton and Simms. These instruments can be fitted with a kinematic stage provided with micrometer adjustments on both the x- and y-motions. It appears to be difficult to make a simple test of the performance of a given stage without special optical devices, but observations on the tracks of the particles with great momentum suggest that, when in good adjustment, any imperfections in the motion of the stage lead to errors which are not greater than those involved in reading the point of intersection of the image of the track and the eye-piece graticule.

The results of a characteristic series of measurements on part of a track of length ·8 mm. is shown in Table I. In this case, values of y have been determined at successive points, 50 μ apart. The observations have been analysed taking a cell length, c, of 100 μ, and the average value of $\bar{\alpha}$ (100 μ) the average angular deviation between successive chords of length 100 μ, is found to be 0·105°. The individual values of D obtained by the above procedure, are not independent, but the use of overlapping cells tends to reduce the effects of reading errors.

A similar analysis has been made using different values of c—200 μ, 300 μ, etc. The mean values of $\bar{\alpha}$ for different values of c are related by the equation $\bar{\alpha}(c_1)/\bar{\alpha}(c_2)=\sqrt{(c_1)}/\sqrt{(c_2)}$. The values of $\bar{\alpha}(200\,\mu)$, $\bar{\alpha}(300\,\mu)$, etc., can therefore be used to determine the value of $\bar{\alpha}(100\,\mu)$ to which they are equivalent, and the corresponding values are included in Table I. It will be seen that $\bar{\alpha}(100\,\mu)$ decreases with increasing cell size and reaches an approximately constant value. This is due to the fact that if the cell-length is too small, the apparent deviations in the track tend to be greater than the true values owing to reading errors, and the resulting value of $\bar{\alpha}$ is too high. On the other hand, if the cell-size is unnecessarily large, the number of statistically independent observations is reduced, and the calculated values of $\bar{\alpha}$ are subject to correspondingly greater statistical fluctuations. It is therefore desirable to employ the smallest possible cell-size which is consistent with the condition that the uncertainty in the value of $\bar{\alpha}$ due to reading errors is small (Goldschmidt-Clermont *et al.* 1948).

TABLE I.

Characteristic results of the measurements of the deviations in the track of a particle of length 8 mm. Readings every 50 μ on first mm. of track. Grain-density in track 19·2 grains/50 μ. $c=100\ \mu$.

x mm.	y dev.	S y_n-y_{n+2}	D S_n-S_{n+2}	x mm.	y dev.	S y_n-y_{n+2}	D S_n-S_{n+2}
12·8	5·6	0·5	0·3	12·25	7·0	0·4	0·3
12·75	5·7	0·6	0·6	12·2	7·3	0·2	0·2
12·7	6·1	0·2	0·1	12·15	7·4	0·1	0·1
12·65	6·3	0·0	−0·2	12·1	7·5	0·0	−0·3
12·6	6·3	0·1	0·0	12·05	7·5	0·0	−0·3
12·55	6·3	0·2	0·1	12·0	7·5	0·3	0·3
12·5	6·4	0·1	−0·2	11·95	7·5	0·3	0·3
12·45	6·5	0·1	−0·3	11·9	7·8	0·0	−0·1
12·4	6·5	0·3	−0·2	11·85	7·8	0·0	−0·1
12·35	6·6	0·4	0·0	11·8	7·8	0·1	0·0
12·3	6·8	0·5	0·3				

Values of $\bar{\alpha}(100\ \mu)$ for different values of c.

c microns	$\overline{\mathrm{D}}$	$\bar{\alpha}(100\ \mu)^{\circ}$
50	0·141	0·191
100	0·22	0·105
150	0·285	0·074
200	0·35	0·060
300	0·54	0·050
400	0·78	0·047
600	1·56	0·051
1000	3·74	0·057
1400	6·28	0·058
2000	11·0	0·058

For small values of c, it is found that, owing to reading errors, values of $\overline{\mathrm{D}}$ tend to become constant, and independent of .c It will be convenient to follow previous authors and refer to this as the "noise-level" ($\overline{\mathrm{D}}_n$). The actual values of $\overline{\mathrm{D}}_n$ depend on a number of factors such as the grain-density in the track, the quality and adjustment of the optical system of the microscope, and the individual skill of the observer, but they are of the order of 0·2 μ.

As the cell-size, c, is increased from these low levels, the value of $(\overline{D})$ rises above the noise level $(\overline{D}_n)$. The corresponding values of $\bar{\alpha}(100\,\mu)$ then rapidly approach a constant value apart from statistical fluctuations. These effects may be summarized by saying that the errors of measurement become unimportant if $\overline{D} > 4\overline{D}_n$. The minimum value of c which gives this condition is adopted in the determination of $\bar{\alpha}(100\,\mu)$. It is clear that if the value of $\bar{\alpha}$ is to have a given statistical weight, the minimum length of track accepted for measurement will depend on the value of $\bar{\alpha}(100\,\mu)$; long tracks being necessary for particles of great energy. Thus if L is the length of the track, the number of statistically independent values of D is $(L/c)-1$. It can be shown that the root-mean-square of the deviations of the individual values of D, from the true mean value corresponding to the momentum of the particle, is equal to $\sim 0{\cdot}5D$. Approximate values of the minimum length of track required to ensure that the probable error in the determination of $\overline{D}$ is less than 25 per cent. are included, for different values of $\bar{\alpha}(100\,\mu)$, in Table II.

Table II.

Values of the minimum length of a track which allows the scattering parameter, $\alpha(100\mu)$, to be determined with a probable error of less than 25 per cent. in the absence of distortion.

Length of track mm.	$\alpha(100\mu)$ degrees
3	0·017
5	0·008
10	0·003
20	0·001

Effects of Distortion of the Emulsion.

When an exposed plate is processed, there is a large increase in the volume of the emulsion. If there is good adhesion between the gelatine and the glass, the swelling in the central regions of the plate corresponds approximately to a uniform linear transformation in the Z co-ordinates for any point ; in the direction, that is, perpendicular to the glass. After fixation and washing, the drying of the plate produces the well-known shrinkage associated with loss of water and the removal of the silver salt from the emulsion during fixation. Since we are concerned only with deviations in the plane projection of a track, such a shrinkage will not, in principle, affect the measurements. In practice, however—and especially with thick emulsions, and in regions near the edges of the plate—some distortion of the emulsion occurs, so that a strictly rectilinear track will, in general, appear to be curved.

The distortion of the emulsion appears to arise, at least, in part, as a result of stresses released during the absorption of water, and because the top surface and sides are unconstrained. An emulsion, of which the cross-section in a plane perpendicular to the glass is rectangular when the emulsion is dry, becomes distorted, therefore, when placed in the processing solutions. Further, the distortion may persist or be increased when the plate is dried. The difficulties can be greatly reduced by using large plates and confining the observations to the central areas ; but even with these precautions, deviations due to distortion can give rise to serious errors in the determination of the energy of the fastest particles. ($\bar{\alpha}(100\,\mu) \lesssim 0{\cdot}01°$ per $100\,\mu$). In the worst cases, corrections may be necessary with tracks for which the true value of $\bar{\alpha}(100\,\mu)$ is as great as $0{\cdot}05°$.

Evidence for the existence of distortion can be obtained in two ways : first, it is sometimes found that the deviations, D, instead of being approximately equally distributed between positive and negative values, are all of one sign. In such cases, an inspection of more steeply dipping tracks, of which the projection is approximately parallel to that of the track under examination, can show whether distortion is indeed present.

Observations show that the distortion is commonly of such a character that the deviation which it produces in a strictly rectilinear track can be written as $\Delta = \mathrm{F}\delta z$, (1), where Δ is the change indirection in the track in passing from a depth z to $z+\delta z$ in the emulsion, and F is a constant for a particular direction of the projection of the track in a given region of the plate. From observations on a more steeply dipping track, the constant F can be determined and appropriate corrections then made to the track under examination. If these corrections change the values of the deviations attributed to scattering so that they are now (apart from differences associated with statistical fluctuations) symmetrically distributed about the value zero, the corrected values are accepted as adequate for the determination of the true value of $\overline{\mathrm{D}}$.

A better method, and the one most commonly adopted, is the following : distortion of the type represented by equation (1) has the result that the apparent deviations in the track are due to the joint effect of the deviations due to scattering, D, and those due to distortion, ϕ. The apparent deviations, D_a, may therefore be written as $\mathrm{D}_a = \mathrm{D} + \phi$, where ϕ is a constant term*. If the distortion is severe, so that ϕ is large, the values of D_a will be all of one sign in spite of the fact that values of D are equally positive and negative apart from statistical fluctuations.

An approximate value of ϕ may be obtained from the quantity $\Sigma\mathrm{D}_a/n \sim \phi$, where $\Sigma\mathrm{D}_a$ represents the algebraic sum of the individual values of D_a. With this value of ϕ, the individual values of D can be found and hence an approximate value of $|\overline{\mathrm{D}}|$. The value of $|\overline{\mathrm{D}}|$ thus obtained is insensitive to the actual value of ϕ employed provided the error in ϕ is less than $|\overline{\mathrm{D}}|/2$.

Table III. shows the results of measurements on a track which shows evidence of severe distortion. In addition to the correction ϕ calculated by the method described above, a number of arbitrarily chosen values have been employed, and the results summarized in Table III. illustrate the fact that the final result is insensitive to the precise magnitude of the correction term.

---

* It is a sufficient approximation to assume that the track "dips" at a constant inclination to the plane of the emulsion.

Table III.

Observation on track showing evidence of distortion.

x mm.	y divs.	S $y_n - y_{n+2}$	D $S_n - S_{n+2}$	D* $\phi=4$ $(D-\phi)$
9·1	10·0	--0·5	5·0	1·0
8·9	9·5	1·0	7·0	3·0
8·7	9·5	4·5	6·0	2·0
8·5	10·5	8·0	4·5	0·5
8·3	14·0	10·25	5·0	1·0
8·1	18·5	12·5	5·25	1·25
7·9	24·25	15·25	4·0	0·0
7·7	31·0	17·75	2·0	−2·0
7·5	39·5	19·25	1·25	−2·75
7·3	48·75	19·25	2·75	−1·25
7·1	58·75	20·5	3·25	−0·75
6·9	68·5	22·5	2·0	−2·0 .
6·7	79·25	23·75		
6·5	91·0	24·5		
6·3	103·0			
6·1	115·5			

$\phi = \Sigma D/n = 4$; $\Sigma D^*/n = 1{\cdot}46$

Values of $\alpha^*(100\,\mu)$ for different values of ϕ.

ϕ	$\Sigma(D-\phi)$	$\alpha^*(100\,\mu)$
0	48·0	0·12
1	36·0	0·09
2	25·5	0·064
3	20·0	0·050
4	17·5	0·044
5	18·5	0·046
6	26·0	0·065
7	36·0	0·090
8	48·0	0·12

Limitations of the Method.

The minimum value of $\overline{D}$ which can be regarded as significant, and therefore the maximum value of the energy of a particle which can be measured, depends on the length of the track available for measurement, the magnitude of the " noise-level ", and the degree of distortion of the gelatine. The maximum measurable values of the energy corresponding to tracks of different length given in Table II., correspond to conditions in which distortion is absent. The minimum value of $\bar{\alpha}(100\,\mu)$ actually observed in the present experiment is 0·001°, and we have no evidence to show whether the theoretical figure of $\bar{\alpha}(100\,\mu)=0{\cdot}0005°$ for a track of length 2 cm., can in fact be attained in our experimental conditions. Such an observation, to be significant, would demand that the motion of the stage departs from being strictly rectilinear by an amount less than $0{\cdot}4\,\mu$ in traversing a distance of 2 cm. That a performance of this order of precision is obtained for shorter lengths of travel is proved by the observation of long tracks with a value of $\bar{\alpha}(100\,\mu)=0{\cdot}001°$.

If distortion exists, then the minimum value of $\bar{\mathrm{D}}$ which can be accepted as significant for a track of a given length is raised above the values given in Table II. By confining the observations to tracks in the central regions of the plates, however, and with the convention that the distortion correction ϕ is less than $2\bar{\mathrm{D}}$, the number of tracks which are rejected when, but for distortion they would have given significant results, is very small.

Calibration.

It has been shown (E. J. Williams 1939, 1940; Rossi and Greisen 1941) that if the track of a particle is divided into cells of length c, the average angular deviation between successive chords $\bar{\alpha}$, is related to the velocity v, of the particle by the equation

$$\bar{\alpha}=\frac{k\mathrm{Z}}{m_0v^2}(1-\beta^2)^{\frac{1}{2}}\left\{\left(\sqrt{\log\frac{\theta_{\max}}{\theta_{\min}}}\right)\right\}, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $\mathrm{Z}e$ is the charge on the particle, and m_0 its rest-mass. The constant in equation (2) depends on the composition of the emulsion, and on the chosen length of cell, c. It is a sufficient approximation to neglect changes in the value of the logarithmic term in (2).

$$\bar{\alpha}=\frac{k_1\mathrm{Z}}{m_0v^2}(1-\beta^2)^{\frac{1}{2}}=k_2\frac{\mathrm{Z}}{pv}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where m_0 is the rest-mass of the particle and p its momentum.

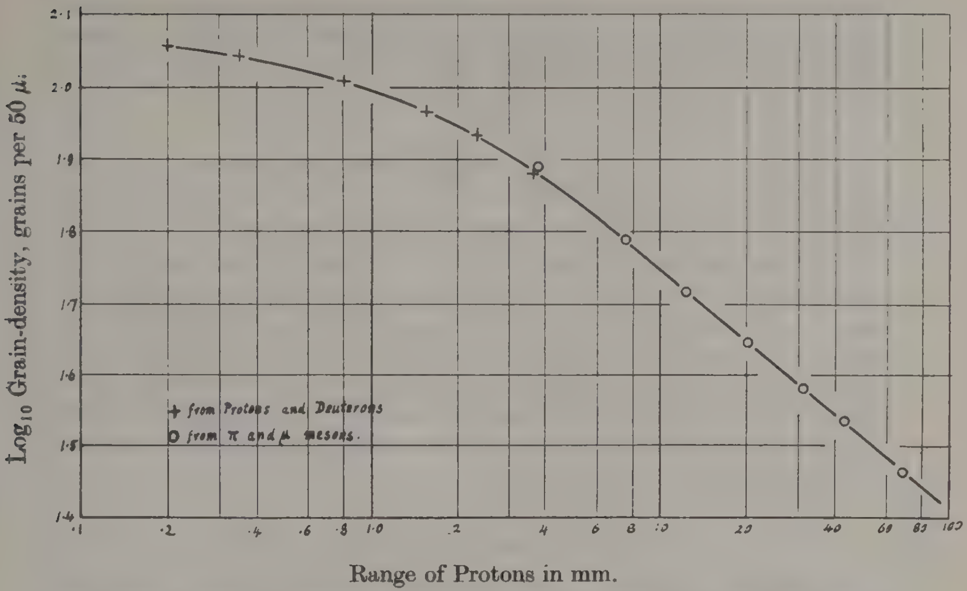
In order to determine the relation between the grain-density, g, in a track, and the rate of loss of energy of the particle producing it, observations have been made on the tracks of mesons, protons and deuterons of great range which stop in the emulsion. The results are shown in fig. 1. This curve, and the range-energy relation for the same particles, would allow us to determine the grain-density in the track of any particle of charge e, as a function of its velocity.

Direct determinations of the range-energy relation for protons in "electron-sensitive" emulsions have not yet been published. It is known, however, that in the Ilford C2 emulsions, the range R, in microns, of a proton with an energy E MeV., is given by the relation

$$\mathrm{R}=\mathrm{CE}^{n}$$

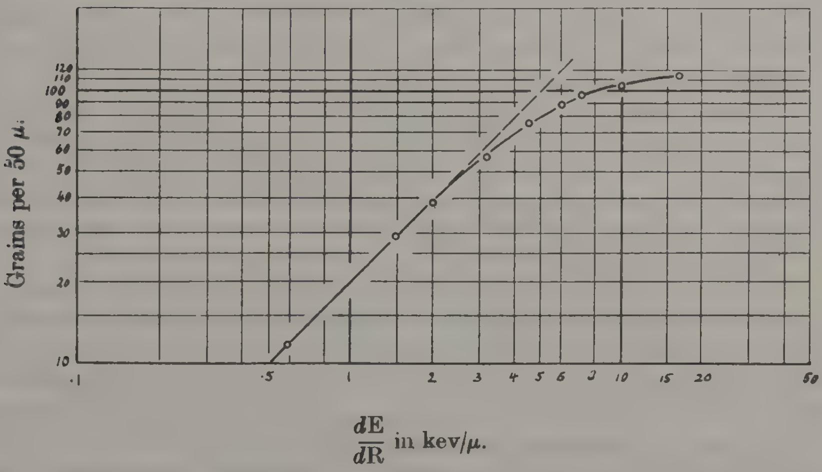
and that $d\mathrm{E}/d\mathrm{R}=kz^2\beta^n$ (4). Assume, therefore, that in the G5 emulsion a relationship of a similar type exists. The corresponding values of k and n can be determined from the following observations :

Fig. 1.



Grain-density in the track of a proton as a function of its residual range.

Fig. 2.



Relation between grain-density and rate of loss of energy along the trajectory. Ilford G.5 emulsion.

(1) The average range in this emulsion of μ-mesons, emitted with a kinetic energy of 4·1 MeV. from a π-meson at rest, is $595 \pm 10\ \mu$.

(2) From fig. 1, we determine the residual range of a proton which has an initial ionization, as determined from the grain density in its track, equal to that of energetic " primary " α-particles with a specific ionization equal to four times the minimum value for charge $|e|$.

From these two observations it is found that

$$\frac{dE}{dR}=\frac{0{\cdot}587\,Z^2}{\beta^{1{\cdot}46}}\,\text{KeV. per micron.} \quad . \quad . \quad (5)$$

This value of n is identical with that for the C2 emulsion.

Using this relation, together with the variation of the grain-density along the track of protons shown in fig. 1, we can determine the grain-density in the track as a function of the rate of loss of energy of the particle; see fig. 2. It will be seen that the relationship is almost linear for values of the rate of loss of energy up to four times the minimum value. From equation (5) and fig. 2, the relation between the grain-density in the track of an energetic proton, and the corresponding value of the scattering parameter $\bar{\alpha}$, can be determined.

EXPERIMENTAL RESULTS.

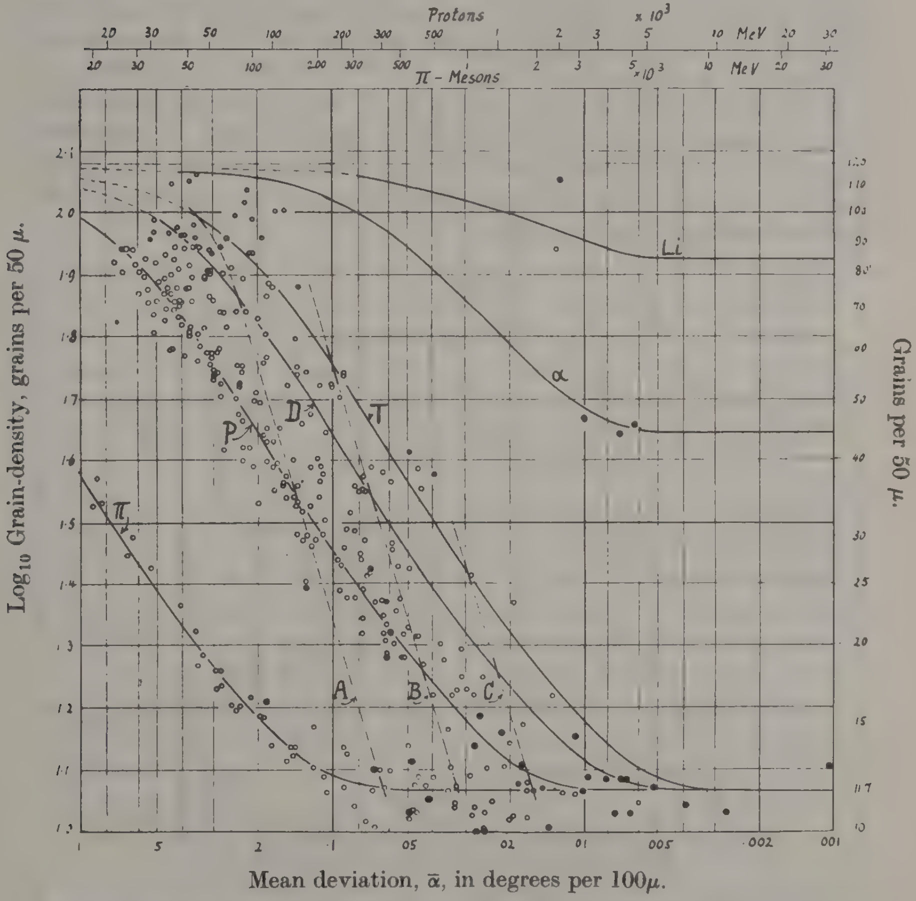
The methods employed in the previous paragraphs have been applied in a study of 350 tracks, associated with " stars ", and chosen only because their length in the emulsion is greater than $3000\,\mu$. The observations were made in plates exposed at high altitudes as described in Part II. of the present series of papers. The results are represented in fig. 3 where the value of the scattering parameter, $\bar{\alpha}$, for a given track, is plotted against the corresponding value of the grain-density, g, both quantities being on a logarithmic scale.

The lines in fig. 3, marked π, P, D, and T are of the form to be expected for particles of charge e as determined by the methods of the preceding paragraphs. They are, however, displaced by about 7 per cent from the calculated curves for the corresponding particles along the axis of the scattering parameter. Their actual positions have been determined by the following method :

The calculated curve for protons has been displaced until it appears to give the best fit with the points corresponding to the most prominent group of particles. These are certainly due to protons. The other curves, with the same form as the first, are then drawn to correspond to particles of mass 286 m_e, for π-mesons ; and with appropriate values of the mass for deuterons and tritons. The corresponding curves for α-particles, lithium nuclei, etc. can also be deduced from the curve for protons using the information given in fig. 2. The difference of only 7 per cent between the calculated curve for protons and the observed distribution for these particles is not unexpected in view of the uncertainty in some of the quantities on which the calculations are based, but it remains to be explained.

In plotting the results shown in fig. 3, primary and secondary particles have been distinguished by the criteria described in (I.); (Brown *et al.* 1949). In the case of events in which many shower particles are emitted, a particle which produces a track with a direction near the axis of the cone

Fig. 3.

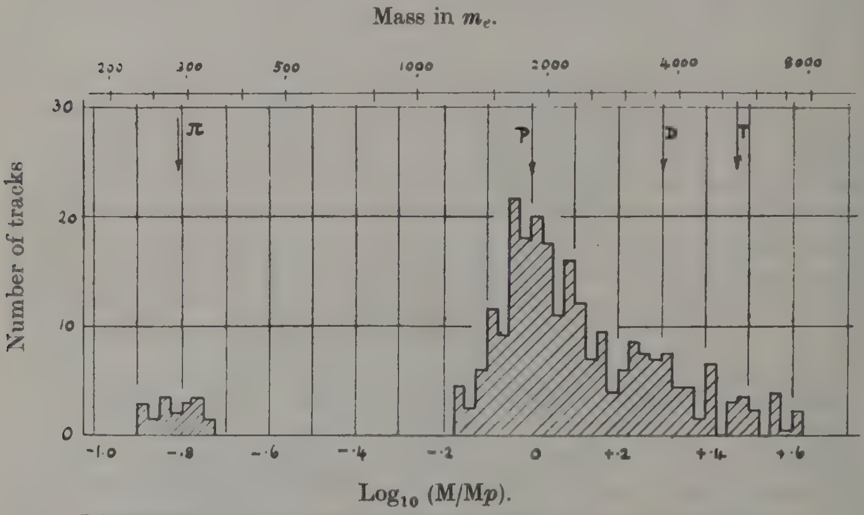


Relation between the grain-density and $\bar{\alpha}(100\mu)$, for the tracks of 350 particles. "Primary" particles are shown thus ●; and "secondary" particles thus ○. The line P has the same form as the calculated curve for protons, but is displaced to the left, in the direction of the $\bar{\alpha}$ axis, by 7 per cent. The lines "π", "D" and "T" are drawn in the correct relation to line P to correspond to particles of mass 286 m_e, 3674 m_e and 5511 m_e. "A", "B" and "C" are lines of constant momentum, and correspond to 500, 1000 and 2000 MeV./c, respectively. Energy scales, in MeV., for π-particles and protons are included at the top of the figure. It will be seen that the α-particles of great energy are all "primaries". The two points corresponding to particles of greater mass are due to a "primary" nitrogen and the "secondary" lithium nucleus occurring in the same event.

embracing the secondary particles, but on the opposite side of the star from the "shower" particles, is regarded as the primary. In the case of events of lower multiplicity, and especially in observations at great altitudes where the directions of motion of the primary particles are more widely dispersed than at 11,000 feet, the identification is subject to some uncertainty, and must be treated with reserve. The greater values of the energy of the primary, as compared with the secondary particles (see fig. 3) strongly suggests, however, that the errors in identification are not serious.

It will be seen, from an inspection of fig. 3, that the primary α-particles produce disintegrations with a frequency of the same order or magnitude as the protons in the same range of energies. On the other hand, compared with protons and deuterons, α-particles of great energy are emitted from nuclear disintegrations very infrequently.

Fig. 4.

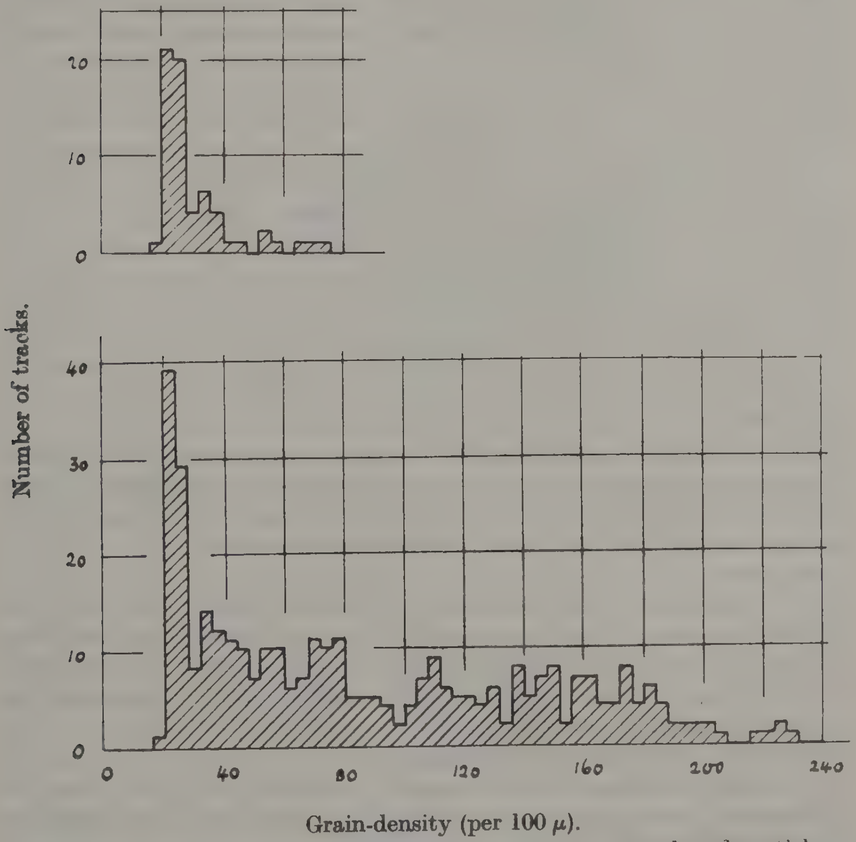


Distribution in mass of the particles of charge |e|, corresponding to the measurements represented in fig. 3.

The displacement from the proton curve, in a direction parallel to the z-axis, of any observed point in fig. 3 gives a measure of the mass of the corresponding particle. The distribution in the values thus obtained—based on measurements of all the tracks represented in fig. 3 which have values of the grain-density lying in the interval from 15 to 98 grains per 50 μ— is shown in fig. 4. It will be seen that in addition to the prominent peak due to protons, there are groups of smaller intensity, unresolved from the protons, which may be attributed to deuterons and tritons; and in addition, a well-defined group due to mesons of which the mean mass is 283 ± 14 m_e.

It may be emphasized that if, as an alternative procedure, the displacement of the individual points from the theoretical curve for protons had been measured, the resulting distribution in the values of the mass of the particles would have been identical in form with that shown in fig. 4, but all the values would have been lowered by 7 per cent. It is reasonable, however, to deduce the mass of the mesons by a direct comparison with that of protons, and the result obtained is identical with that given above. The majority of the mesons may therefore be identified as π-particles

Fig. 5.



Distribution in the values of the grain-density of the tracks of particles of charge |e|. The lower distribution is for all particles ; and the upper, for those which are identified as π-mesons.

The observed distribution in the values of the grain-density of all the tracks of particles of charge e represented in fig. 3 is shown in fig. 5. This curve may be compared with the results of similar measurements made on plates exposed on Jungfraujoch and given in (I.), fig. 1. The two distributions are similar in form, but there are relatively more particles of low grain-density in the present observations. This corresponds to the greater average energy of the nuclear explosions at the greater altitude, see (II.).

A study of fig. 3 shows that most of the particles producing tracks with a grain-density less than 14 grains per 50 μ are either mesons or " primary " particles. The nature of the secondary particles for which $\bar{\alpha}(100\ \mu)<0{\cdot}02^\circ$ cannot be determined by our present methods. It is possible, however, to extrapolate the observed distributions in energy of protons and π-mesons, in the range in which they can be distinguished, to the high energy region, and thus to estimate the proportion of the two types. Making suitable corrections for the " escape " of particles from the emulsion, the relative frequency of occurrence of the different types can be estimated, and the results are shown in Table IV.

TABLE IV.

Number of particles of different types, per " star ", with a range greater than 3 mm., observed at an altitude of 70,000 feet.

π-particles	0·6,	Protons	1·4,
Deuterons	0·5,	Tritons	0·2,
α-particles	0·02.		

Average number of tracks per " star "=7·4.

CONCLUSIONS.

The following conclusions can be drawn from the results of the present experiments :

(1) Of the mesons emerging with a kinetic energy less than 150 MeV. from nuclear disintegrations produced by protons and α-particles of great energy, the great majority are π-particles. If μ-particles are sometimes directly created, they are present in numbers less than 2 per cent of the π-particles.

(2) If electrons, or other charged particles of small rest-mass, are sometimes emitted from nuclear explosions, with an energy less than 150 MeV., they occur with a frequency less than 2 per cent of the π-particles. Most of the electrons of the soft component of the cosmic radiation have an energy less than 200 MeV. The observations therefore prove that most of them are of secondary origin.

(3) The observations give no support for the view that mesons of intermediate mass, τ-mesons with a lifetime greater than 10^{-12} sec., are ejected with an energy less than $\sim$300 MeV., during nuclear explosions ; and they prove that if such particles, with a mass between 350 and 1200 m_e, are sometimes created in these processes, their number is less than 2 per cent of that of the π-mesons and less than 1 per cent of the protons. This conclusion is complementary to that drawn recently by Franzinetti (1950) from determinations by magnetic deflection experiments of the mass of slow charged particles produced by cosmic radiation.

(4) No final conclusion can be drawn regarding the nature of the particles, less massive than the proton, which produce tracks with the minimum value of the grain-density. In view of the observations on particles of lower energy, it is reasonable to suppose that the faster particles are of the same type. In what follows they will therefore be regarded as π-particles.

(5) At least 90 per cent of the "shower" particles emerging from nuclear explosions are π-mesons, a "shower" particle being defined as one with a grain-density less than $1{\cdot}5\ g_{\min}$ (see I.). Consider, however, all particles in a given range of momentum—say between 200 and 700 MeV./c—a practice commonly adopted in experiments with Wilson chambers operated in magnetic fields. Among the particles in this momentum range, the π-mesons are accompanied by large numbers of protons. If the π-mesons are equally distributed between positive and negative particles, the "positive excess" should be of the order of six, a value consistent with that observed in experiments with expansion chambers (Rochester 1949).

(6) An extension of the observations will allow a detailed analysis of the nuclear processes produced by protons and α-particles in the range of energies above 10^9 eV., including the distribution in energy of the π-mesons near their points of production, and the relation between the energy of a π-particle and its direction of emission with respect to the "primary" particle. Such investigations are being undertaken in a collaboration with U. Camerini, W. O. Lock and H. Muirhead; and by J. H. Davies and C. Franzinetti.

(7) The observed homogeneity in the values of the mass of the ejected mesons with energy less than 150 MeV. allows us to exclude certain assumptions regarding the origin of the particles. It is known, from the experiments on the artificial creation of π-mesons made at Berkeley (Gardner and Lattes 1949), that the threshold for the process occurs at such an energy of the bombarding protons or α-particles that the creation of a particle of mass $\sim 280\ m_e$ is just possible from the point of view of the conservation laws. It follows that the π-mesons must be directly created; that they cannot be regarded as the products of decay of mesons of appreciably greater mass, and very short lifetime.

In the nuclear transmutations with which we are concerned, limitations of energy do not apply. In principle, therefore, the possibility cannot be excluded that the π-mesons occurring in showers are produced by the decay of short-lived mesons of greater mass which are the primary products of the nucleon–nucleon collisions. The observed homogeneity in the values of the mass of the ejected mesons proves, however, that any postulated unstable parent of the π-particles, if it exists, must be very short-lived; for if it survived, on the average, over a length of path of say 2 mm., the observed values of the mass would represent averages between the masses of the parent and daughter particles. As a result, we should expect to observe a dispersed set of values instead of the compact group actually found. The present observations therefore prove that any more massive parent mesons have lifetimes less than 10^{-13} sec. It is, of course, reasonable to suppose that the π-mesons are directly produced in the nucleon–nucleon collisions.

ACKNOWLEDGMENTS.

In conclusion, the author wishes to record his sincere thanks to Professor C. F. Powell for his encouragement and help; to Dr. H. Heitler and Mr. J. Webb for their assistance; to Messrs. W. O. Lock and H. Muirhead for processing the Ilford G. 5 plates, with emulsions 400μ thick, employed in the investigation; and to the team of observers in this Laboratory who scanned the plates.

REFERENCES.

BROWN, CAMERINI, FOWLER, HEITLER, KING and POWELL, 1949, *Phil. Mag.*, **40,** 862.

CAMERINI, COOR, DAVIES, FOWLER, LOCK, MUIRHEAD, and TOBIN, 1949, *Phil. Mag.*, **40,** 1073.

DAVIES, LOCK, and MUIRHEAD, 1949, *Phil. Mag.*, **40,** 1250.

FRANZINETTI, 1950, *Phil. Mag.*, **41,** 86.

GOLDSCHMIDT-CLERMONT, KING, MUIRHEAD, and RITSON, 1948, *Proc. Phys. Soc.*, **61,** 183.

ROSSI, and GREISEN, 1941, *Rev. Mod. Phys.*, **13,** 240.

ROCHESTER, 1949, *Report to Como Conference.*

WILLIAMS, 1939, *Proc. Roy. Soc. A*, **169,** 531;

—— 1940, *Phys. Rev.*, **58,** 292.

XVI. *The S-Matrix for Meson-Nucleon Interactions.*

By P. T. Matthews,
Clare College, Cambridge, England *.

[Received Nov. 14, 1949.]

Abstract.

The possibility is considered of obtaining a divergence free S-matrix for the interactions of mesons with nucleons along the lines of mass and charge renormalization, which Dyson has so successfully applied to quantum electrodynamics. It has been shown that this can be done for the vector interaction of neutral vector mesons. The procedure is here extended to the scalar interactions of scalar and pseudo-scalar mesons but proves to be quite inadequate to remove all the divergences in the other cases.

§ 1. Introduction.

By applying Feynman's (1949 a, b) graphical methods for the calculation of matrix elements to relativistic quantum electrodynamics in the Schwinger (1948) form, Dyson (1949 a, b) † has shown how a finite S-matrix can be obtained for a wide class of phenomena. In Feynman's theory, any real process taking place with specific virtual processes can be represented by a graph from which the integral for the corresponding matrix element can be written down directly, each line and each vertex corresponding to a single factor in the integrand. An essential feature of Dyson's development is the proof that only a limited number of different types of divergence can occur in quantum electrodynamics. These are called primitive divergents. When a primitive divergent graph occurs as part of a larger graph its effect is to complicate a single line or a single vertex of a simpler graph, and consequently to modify a single factor of the integrand corresponding to the simpler graph. If the infinite parts of all such modifications of a graph are dropped in a well-defined way, a finite S-matrix is obtained for all phenomena which do not involve bound states. The mass and charge of a free particle in the vacuum are both given by the theory as infinite renormalizations of the mass and charge of an ideal 'bare' particle. Dyson has shown that his prescription for dropping the infinite terms is equivalent to equating the infinite theoretical values of mass and charge to the finite experimental values.

The extension of this method to the various meson-nucleon interactions is here considered. The number and degree of the possible primitive divergents in any theory are closely related to the degree of the derivatives of Δ-functions which occur in the commutation relations of the Hamiltonian under consideration. The meson interactions which in this respect most closely resemble electrodynamics are the scalar interactions of scalar and pseudo-scalar mesons with nucleons in which no derivatives occur from meson variables. The possible types of primitive divergents are here identical with those of electrodynamics. In electrodynamics one of these —the scattering of light by light—is finite by gauge invariance, but the scattering of mesons by mesons gives rise to genuine divergences. However, it will be shown that these can be regarded as a special type of self-energy between pairs of mesons and can be cancelled by adding the appropriate δ-type interaction between mesons to the Lagrangian. A finite S-matrix will thus be obtained. In the remaining meson-nucleon interactions double derivatives arise from the commutation relations of the meson variables with the result that there is no limit to the number and degree of the possible types of primitive divergents. Thus the renormalization of mass and charge is shown to be quite inadequate for the removal of all divergences in these cases. The only exception to this is the vector interaction of neutral vector mesons with nucleons, which has already been dealt with elsewhere (Matthews 1949 b).

---

* Communicated by Dr. N. Kemmer.

† These papers will be referred to as I. and II.

The proofs of this paper do not confirm exactly the conclusions of Case (1949) which were generalizations from the results of particular calculations. Some of the infinities considered here only occur in comparatively high order corrections and were thus not detected.

§ 2. THE S-MATRIX.

The interaction of two fields can be expressed by a generalized Schrödinger equation in the interaction representation,

$$i\hbar c\delta\Psi[\sigma]/\delta\sigma(x)=\mathrm{H}(\phi(x))\Psi[\sigma], \quad \ldots \ldots \ldots \quad (1)$$

where $\Psi[\sigma]$ is the wave function on a general space-like surface, σ, and $\phi(x)$ is a typical field variable satisfying the equation of motion and the commutation relations of its free field with mechanical mass. The Hamiltonian in (1) has matrix elements for transitions which cannot occur with conservation of energy and momentum, and it is convenient to remove these by a Schwinger (1948) transformation

$$\Psi\rightarrow\exp\{-i\mathrm{S}(\phi)\}\Psi. \quad \ldots \ldots \ldots \quad (2)$$

The new equation is

$$i\hbar c(\delta/\delta\sigma)(\exp\{-i\mathrm{S}(\phi)\}\Psi)=\mathrm{H}(\phi)\exp\{-i\mathrm{S}(\phi)\}\Psi. \quad \ldots \quad (3)$$

To second order in the coupling constant the new Hamiltonian has matrix elements for real processes and also some infinite elements for the self-energies of the interacting particles. If the infinite self-energy parts of the Hamiltonian in (3) are $\mathrm{H}_c(\phi)$, it has been shown for a fermion field ψ and a boson field ϕ_μ, (Matthews 1949 c), that

$$\mathrm{H}_e(\phi,\psi)=\hbar c\delta\kappa_0\bar{\psi}\psi+\tfrac{1}{2}\delta\kappa^2\phi_\mu, \quad \ldots \ldots \ldots \quad (4)$$

where κ_0 and κ are the reciprocal Compton wavelengths for the fermions and the bosons respectively. To obtain an equation from which these terms have been removed make the transformation

$$\Psi = \mathrm{U}\Psi_e, \qquad (5)$$

where

$$i\hbar c\,\delta\mathrm{U}/\delta\sigma = \mathrm{H}_e(\phi)\mathrm{U}. \qquad (6)$$

It then follows that

$$i\hbar c\,\mathrm{U}^{-1}\delta\mathrm{U}/\delta\sigma = \mathrm{H}_e(\phi_e), \qquad (7)$$

where

$$\phi_e = \mathrm{U}^{-1}\phi\mathrm{U}. \qquad (8)$$

The equation which Dyson makes the starting-point of his theory is obtained by reversing the order of these transformations. The transformation (5) on equation (1) leads to

$$i\hbar c(\delta\Psi_e/\delta\sigma) + i\hbar c\,\mathrm{U}^{-1}(\delta\mathrm{U}/\delta\sigma)\Psi_e = \mathrm{H}(\phi_e)\Psi_e.$$

By (7),

$$\begin{aligned} i\hbar c\,\delta\Psi_e/\delta\sigma &= \{\mathrm{H}(\phi_e) - \mathrm{H}_e(\phi_e)\}\Psi_e \\ &= \mathrm{H}_1(\phi_e)\Psi_e, \end{aligned} \qquad (9)$$

which is the Schrödinger equation in the interaction representation in a form which is free from self-energy effects up to the second order in the coupling constant. Since ϕ satisfies the free field equation with mechanical mass, it follows from (4), (6) and (8) that ϕ_e satisfies the free field equation with renormalized mass. This is given infinite by the theory, but since it is the mass of the observed free particle, it is equated to the finite experimental mass. An equation like (3) with the self-energy explicitly subtracted could now be obtained from (9) by performing the transformation (2) with ϕ replaced by ϕ_e.

Dyson's theory starts with an equation of the form (9), but the constants $\delta\kappa_0$ and $\delta\kappa^2$ are chosen as power series in the coupling constant so that they will cancel all the self-energy effects from the rest of the Hamiltonian up to any order. If Ψ_0 is the state vector of a system on a surface in the remote past and Ψ is the same system, after its particles have interacted, in the distant future, then

$$\Psi = \mathrm{S}(\infty)\Psi_0, \qquad (10)$$

where in Dyson's notation (I. § 5; II. § 2)

$$\mathrm{S}(\infty) = \sum_{n=0}^{\infty}\left(\frac{-i}{\hbar c}\right)^n \frac{1}{n!}\int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n\, \mathrm{P}(\mathrm{H}_1(x_1), \mathrm{H}_1(x_2), \ldots, \mathrm{H}_1(x_n)). \qquad (11)$$

$\mathrm{S}(\infty)$ is the Heisenberg S-matrix, and, by developing Feynman's (1949 a, b) methods, Dyson has given comparatively simple rules for the calculation of its elements (I. §§ 6, 7; II. § 2). Essentially, if H_1 contains factors ψ, $\bar{\psi}$ and ϕ, where

$$[\phi(x), \phi(y)] = i\hbar c\Delta(x-y), \qquad (12)$$

and

$$\{\psi(x), \bar{\psi}(y)\} = -i(\gamma_\mu\partial/\partial x_\mu - \kappa_0)\Delta(x-y). \qquad (13)$$

the matrix element for a process in which a virtual particle passes between x and y, contains in the integrand a factor

$$\langle \mathrm{P}(\phi(x), \phi(y))\rangle_0 = \tfrac{1}{2}\hbar c \Delta_{\mathrm{F}}(x-y) \quad . \; . \; . \; . \; . \quad (14)$$

if the particle is a boson, and

$$\langle \mathrm{P}(\bar{\psi}(x), \psi(y))\rangle_0 = -\tfrac{1}{2}\eta(x, y)(\gamma_\mu \partial/\partial x_\mu + \kappa_0)\Delta_{\mathrm{F}}(x-y) = -\tfrac{1}{2}\eta(x, y)\mathrm{S}_{\mathrm{F}}(x-y)$$

if it is a fermion. If, in place of ϕ, H_1 contains a factor $\partial\phi/\partial x_\mu$, then it will also contain a term depending explicitly on the normal to the surface σ. It has already been shown (Matthews 1949 a) that to extend Dyson's rules to this case the normal-dependent terms in H_1 can be neglected and then effectively

$$\langle \mathrm{P}(\partial\phi/\partial x_\mu, \partial\phi/\partial y_\nu)\rangle_0 = \tfrac{1}{2}\hbar c(\partial^2/\partial x_\mu \partial y_\nu)\Delta_{\mathrm{F}}(x-y). \quad . \; . \; . \quad (15)$$

A similar result holds if the derivatives occur in the commutation relations themselves. Thus for vector mesons

$$[\phi_\mu(x), \phi_\nu(y)] = i\hbar c(\delta_{\mu\nu} + (1/\kappa^2)\partial^2/\partial x_\mu \partial y_\nu)\Delta(x-y), \quad . \; . \quad (16)$$

and to calculate terms of (11) for a Hamiltonian containing a factor ϕ_μ, the normal-dependent terms can be neglected and effectively

$$\langle \mathrm{P}(\phi_\mu(x), \phi_\nu(y))\rangle_0 = \tfrac{1}{2}\hbar c(\delta_{\mu\nu} + (1/\kappa^2)\partial^2/\partial x_\mu \partial y_\nu)\Delta_{\mathrm{F}}(x-y). \quad . \quad (17)$$

The infinities in $\mathrm{S}(\infty)$ associated with mass renormalizations have already, in principle, been removed. There remain other infinities which in electrodynamics can be removed entirely by renormalization of the charge. Before considering the possible application of this method to meson interactions, it will be necessary to discuss in more detail Dyson's analysis of the possible types of primitive divergents, because it is at this point that the difficulties of extending the theory arise.

§ 3. PRIMITIVE DIVERGENTS.

Consider an interaction between fermions and bosons in which no derivatives occur either in the interaction or in the commutation relations of the bosons. Every element of $\mathrm{S}(\infty)$ is associated with certain graphs G, the process under consideration determining the number of external fermion lines, E_f, and the number of external boson lines, E_b. Consider any such graph, G, with n vertices and F internal lines of which F_f are fermion lines and F_b are boson lines. Now

$$\Delta_{\mathrm{F}}(x) = \frac{-i}{(2\pi)^3}\int \frac{\exp\{-ip_\mu x_u\}}{p_\mu{}^2 + \kappa^2}\, d^4 p_\mu. \quad . \; . \; . \; . \quad (18)$$

The integrals over the space variables in the elements of $\mathrm{S}(\infty)$ are trivial, giving conservation of energy and momentum at each vertex. Thus the energy-momentum associated with each line in G can be expressed in terms of $(\mathrm{F}-n+1)$ basic vectors p^i, and the corresponding contribution to $\mathrm{S}(\infty)$ is an integral, M, over these variables. For considerations of convergence, each boson line contributes to the integrand of M a factor $1/((\Sigma a_i p^i)^2+\kappa^2)$ and each fermion line a factor $\Sigma a_i p^i_\mu/((\Sigma a_i p^i)^2+\kappa_0^2)$, where $\Sigma a_i p^i$ is the energy-momentum associated with the line. The numerator of the fermion factor comes from the derivative of the Δ-function in the fermion commutation relation (13). Many of the coefficients a_i will be zero and, if for a certain line only a_i is non-zero, the line will be referred to as a basic line. If an internal line in a graph is replaced by two external lines, the line is said to be opened.

The differentials d^4p_μ can be expressed in polar variables. The integrals over the angle variables are finite and the integral over the radius is p^3dp. Define the convergence function

$$K=2F-F_f-4(F-n+1). \qquad (19)$$

If there is only one basic p^i, $-(K+1)$ is the power of p in the integrand of M expressed in polar variables. The condition for convergence is in this case

$$K\geqslant 1. \qquad (20)$$

If there are several basic variables, define K_i to be K evaluated for the graph obtained by opening all the basic lines except p^i, that is for the integral M with all variables except p^i held constant. Similarly define K_{ij} to be K evaluated for the graph with all basic lines except p^i and p^j opened. A sufficient condition for M to be conditionally convergent is that

$$K_i\geqslant 1,\quad K_{ij}\geqslant 1,\ldots, K\geqslant 1, \qquad (21)$$

for all $i, j, \ldots=1, 2, \ldots, (F-n+1)$. The graph is absolutely convergent if (21) holds for all possible choices of the basic variables p^i. This condition is not necessary for convergence because it may happen that the coefficients of the higher powers of p in M vanish identically.

A graph is said to be primitive if the graphs obtained from it by opening any line are convergent. Thus a primitive graph certainly satisfies all the conditions (21) except possibly the last one. If this is also satisfied, the graph is itself convergent. The only possible primitive divergent graphs are those for which the last condition is not satisfied. That is if

$$K>1. \qquad (22)$$

If G has n_f vertices with just two fermion lines incident, and n_b vertices with just two boson lines incident, then

$$2F=3n-n_f-n_b-E_f-E_b, \qquad (23)$$

and

$$F_f=n-n_b-\tfrac{1}{2}E_f. \qquad (24)$$

Thus

$$K=(3/2)E_f+E_b+n_f+2n_b-4, \qquad (25)$$

and the only possible primitive divergents are (i) $E_f=2$, $E_b=0$; (ii) $E_f=2$, $E_b=1$; (iii) $E_f=0$, $E_b=2$; (iv) $E_f=0$, $E_b=4$, and two other graphs which contain odd fermion loops which cancel with the similar graphs obtained by reversing the directions of the lines in the loop. The graphs (i), (ii) and (iii) are self-energy and vertex parts which can be regarded as modifications of single lines or single vertices. The graphs (iv) represent the scattering of bosons by bosons and will be called "square" parts. The simplest, containing four free boson and a loop of four fermion lines, will be called a simple square part, S_1.

Now any divergent graph must be a divergent primitive or contain divergent primitive parts and a convergent graph can only be made divergent by the insertion of primitive divergent parts. Dyson showed, (II. § 6), by general arguments of gauge invariance that for electrodynamics no new infinities arise from the insertion of square parts. These arguments are of course not applicable to mesons, and by Dyson's methods it can be shown by direct calculation that the scattering of scalar or pseudo-scalar mesons by themselves is not finite like the scattering of light by light. Thus for meson interactions a special consideration of square parts is necessary.

For a single free meson in the vacuum the self-energy arises from the virtual creation of nucleon pairs. When there are two free mesons these can scatter each other by virtual nucleon pair creation, (square graph). The matrix element for this process is infinite, but a finite part can be separated off if, in the spirit of Dyson's work, we subtract from it that part which is obtained by setting the energy-momenta of the free mesons equal to zero. This subtracted part can be regarded as a special type of self-energy, which will have the form $\hbar c\delta\lambda\phi^4$, where $\delta\lambda$ is an infinite constant. In the next section, it will be shown how a finite S-matrix can be obtained up to any order for certain meson interactions by the addition of a term of this type to the Lagrangian for the interaction.

Note that the $\delta\lambda$ effect, unlike the $\delta\kappa^2$, is essentially an interaction between mesons which will not take place for a single free meson. The term is thus properly included in the interaction Lagrangian, when it will occur as an additional term in the Hamiltonian of (1) and, after the transformation (5), in the Hamiltonian $H_1(\phi_c)$ of (9). It would not be a reasonable procedure to start with the usual Lagrangian and to add the term to the self-energy Hamiltonian, H_c, of (4). It would then appear both in $H_1(\phi_c)$ of (9) and in the free field Hamiltonian determining the motion of ϕ_c. But the equation of motion of ϕ_c should be the wave equation for a single free meson which, as pointed out above, should not depend on $\delta\lambda$.

§ 4. SCALAR MESON SCALAR INTERACTIONS.

The only interactions, in addition to electrodynamics, which satisfy the condition at the beginning of the last section are the scalar interactions of scalar and pseudo-scalar mesons with nucleons. If the term $\hbar c\delta\lambda\phi^4$ is added to the Lagrangian in the pseudo-scalar case, the Hamiltonian $H_1(x)$ of (9) is

$$H_1(x)=if\bar{\psi}(x)\gamma_5\psi(x)\phi(x)-\hbar c\delta\kappa_0\bar{\psi}(x)\psi(x)-\tfrac{1}{2}\delta\kappa^2\phi^2(x)-\hbar c\delta\lambda\phi^4(x). \qquad (26)$$

This Hamiltonian will give rise to graphs in which the most important vertices will have two nucleon and one meson line incident, at which the first term of (26) is operating. There will also be vertices at which either two nucleon, two meson or four meson lines are incident. At these vertices the terms $\delta\kappa_0$, $\delta\kappa$ and $\delta\lambda$ respectively will be operating. The rules for obtaining the matrix elements from the graphs are the same as those given by Dyson, (II. §§ 2, 3), with the following substitutions. $A_\mu(x)$ is replaced by $\phi(x)$, e is replaced by $-f$, γ_μ is replaced by γ_5 and $D_F(x)$ is replaced by $\Delta_F(x)$. Also if n_l is the number of vertices at which $\delta\lambda$ is operating, in place of II. Eq. (12), terms must be multiplied by

$$(-1)^{n-n_m-n_l-l-m}. \qquad (27)$$

This term makes allowance for the changes of sign which are due to the rearrangement of the ψ and $\bar{\psi}$ factors to bring associated pairs adjacent, (I. §§ 6, 7). As in electrodynamics, functions $\Sigma(W)$, $\Pi(W')$ and $\Lambda(V)$ can be defined, (II. Eqs. (27), (39), (32)) through which can be determined the alteration in the integrand due to the insertion of a nucleon self-energy part, W, a meson self-energy part, W′, or a vertex part, V. If W consists of a single vertex,

$$\Sigma(W, t^1) = -2\pi i \delta\kappa_0. \qquad (28)$$

If W′ is a single vertex,

$$\Pi(W', t^1) = \pi i \delta\kappa^2. \qquad (29)$$

The different signs of the terms in (28) and (29) are due to the factor (27). It is also necessary to consider explicitly the integrals from square graphs, S, which will be of the form

$$\Delta_F(t^1)\Delta_F(t^2)\Delta_F(t^3)\Delta_F(t^1+t^2-t^3)K(S, t^1, t^2, t^3).$$

In general, K(S) is itself a function of Δ_F, S_F and γ_5. In the special case when S is a single vertex,

$$K(S, t^1, t^2, t^3) = i\delta\lambda. \qquad (30)$$

In Dyson's terminology, a " proper " vertex or self-energy part is one which cannot be split into two parts joined by a single line. An " irreducible " graph is one from which all self-energy and vertex parts have been removed. It is also convenient to consider irreducible self-energy and vertex parts, which are such parts that are proper and from which all internal self-energy and vertex parts have been removed. We now extend these definitions. A " quasi-proper " square part is a square part which cannot be split into a square part joined to a self-energy part by a single line. A " proper " square part is a quasi-proper square part which cannot be split into two square parts joined by two meson lines. " Totally irreducible " graphs are irreducible graphs which contain only simple square parts. $\Sigma(t^1)$ is the sum of $\Sigma(W, t^1)$ over all fermion self-energy parts, and $\Sigma^*(t^1)$ is the sum over all proper fermion self-energy parts. $\Pi(t^1)$ and $\Pi^*(t^1)$ are similarly defined. Square parts which are not quasi-proper are clearly redundant. Thus $K(t^1, t^2, t^3)$ is defined as the sum over all quasi-proper square parts and $K^*(t^1, t^2, t^3)$ as the sum over all proper square parts.

To obtain the matrix element for a given process, one must consider the totally irreducible graphs for the process. The corresponding integrals will be made up of factors S_F, Δ_F, γ_5 and $K(S_1)$ (from the simple square parts). To make allowance for the insertion of self-energy parts into the lines of such graphs, the factors S_F and Δ_F must be replaced by S'_F and Δ'_F, which are implicitly defined by the equations (II. Eqs. (63), (64)),

$$S'_F(t^1)=S_F(t^1)+S_F(t^1)\Sigma^*(t^1)S'_F(t^1), \quad . \quad . \quad . \quad . \quad (31)$$

$$\Delta'_F(t^1)=\Delta_F(t^1)+\Delta_F(t^1)\Pi^*(t^1)\Delta'_F(t^1). \quad . \quad . \quad . \quad . \quad (32)$$

These equations express the fact that a self-energy part is either proper or is a proper self-energy part attached by a single line to a self-energy part, which may be proper or improper. The insertion of all quasi-proper square parts in place of a simple square part is allowed for by replacing the factor $K(S_1)$ by K. Now a quasi-proper square part is either a proper square part or is a proper square part joined to a quasi-proper square part by two meson lines, which may themselves contain self-energy parts. Thus, corresponding to (31) and (32), K is implicitly defined by the equation

$$K(t^1, t^2, t^3)=K^*(t^1, t^2, t^3)$$
$$+\int K^*(t^1, t^2, t)\Delta'_F(t)\Delta'_F(t^1+t^2-t)K(t, t^1+t^2-t, t^3)\, d^4t. \quad . \quad (33)$$

No infinity will arise from the integral over t. $\Lambda(t^1, t^2)$ is summed over all proper vertex parts only, and the effect of inserting vertex parts into the totally irreducible graphs is allowed for by replacing γ_5 by $\Gamma(t^1, t^2)$, which is defined by

$$\Gamma(t^1, t^2)=\gamma_5+\Lambda(t^1, t^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

All the infinities in the matrix elements will appear in the functions S'_F, Δ'_F, Γ and K. It must now be shown how these can be removed.

Consider for the moment only totally irreducible parts. If S is a totally irreducible square part, then $K(S, t^1, t^2, t^3)$ contains divergent factors $K(S_1)$ from the simple square parts contained in S. These simple square parts can each be replaced by a single vertex at which $\delta\lambda$ is operating, and the first term in the expansion of $\delta\lambda$ (of order f^4) can be chosen to cancel these infinities. (The higher-order terms in $\delta\lambda$ will be defined later to cancel certain other infinities which arise when irreducible square parts are considered.) With this understanding, the integral $K(S, t^1, t^2, t^3)$, as a whole, still diverges logarithmically. By splitting off the part obtained by setting $t^1=t^2=t^3=0$ and summing over all totally irreducible square parts,

$$K^*(t^1, t^2, t^3)=A''+i\delta\lambda+K^*_c(t^1, t^2, t^3), \quad . \quad . \quad . \quad . \quad (35)$$

where A'' is an infinite constant and K^*_c is a uniquely defined finite function. With the same understanding about simple square parts, the logarithmically divergent $\Lambda(V, t^1, t^2)$ can be split into a finite constant and a finite part. By summing over all totally irreducible vertex parts, V, we have

$$\Lambda(t^1, t^2)=L\gamma_5+\Lambda_c(t^1, t^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

Totally irreducible self-energy parts do not contain any square parts (see below), and can be split up as in electrodynamics (II. Eqs. (67, (77)),

$$\Sigma^*(t^1)=\mathrm{A}-2\pi i\delta\kappa_0+\mathrm{B}(t^1_\mu\gamma_\mu-i\kappa_0)+(t^1_\mu\gamma_\mu-i\kappa_0)\mathrm{S}_c(t^1), \quad . \quad . \quad (37)$$

$$\Pi^*(t^1)=\mathrm{A}'+\pi i\delta\kappa^2+\mathrm{C}((t^1)^2+\kappa^2)+((t^1)^2+\kappa^2)\Delta_c(t^1). \quad . \quad . \quad (38)$$

L, A, B, A′ and C are infinite constants and $\Delta_c(t^1)$, $\mathrm{S}_c(t^1)$ and $\Lambda_c(t^1, t^2)$ are defined by absolutely convergent integrals. Substituting these expressions into (31), (32), (33) and (34), equations are obtained for $\mathrm{S}'_{\mathrm{F}}(t^1)$, $\Delta'_{\mathrm{F}}(t^1)$, $\mathrm{K}(t^1, t^2, t^3)$ and $\Gamma(t^1, t^2)$. $\mathrm{S}'_{\mathrm{F}1}$, $\Delta'_{\mathrm{F}1}$, Γ_1 and K_1 are defined to be the functions which are obtained if the infinite constants are simply dropped from these equations. Note that so far no account has been taken of self-energy, vertex and square parts which are not totally irreducible.

It will now be shown that, by successive approximation, it is possible to develop power series in the coupling constant for finite functions $\mathrm{S}'_{\mathrm{F}1}$, $\Delta'_{\mathrm{F}1}$, K_1 and Γ_1 even when graphs which are not totally irreducible are considered. Suppose that the series for these functions have been developed up to f^{2n}. To develop the next term of the series for K^*_c, consider the contribution from a totally irreducible square graph, S, of order l, ($l \leqslant 2n+4$). The integral corresponding to this graph will contain factors S_{F}, Δ_{F} and γ_5 and factors $\mathrm{K}(\mathrm{S}_1)$ for the simple square parts. The exact integral is obtained by replacing these by $\mathrm{S}'_{\mathrm{F}1}$, $\Delta'_{\mathrm{F}1}$, Γ_1 and K_1 respectively. These will not introduce any divergencies and will determine the integrand up to the order f^{2n+l}. The whole integral will itself diverge logarithmically, but the finite part can be separated from an infinite constant by the usual procedure to determine $\mathrm{K}_c(\mathrm{S})$ and $\mathrm{A}''(\mathrm{S})$ to the order f^{2n+l}. Since $l \geqslant 4$, by summing over all such graphs, S, power series in f can be developed for K_1 and A″ up to the order f^{2n+4}.

To obtain Γ_1 to order f^{2n+2}, sum in the same way over all totally irreducible vertex graphs, V., after substituting the corrected functions for the line, vertex and simple square factors. The logarithmically divergent integrals can again be split into an infinite constant and a convergent part to give power series for L and Λ_c up to f^{2n+2}. Hence Γ can be found to the same order and Γ_1 by dropping the infinite constant L.

To obtain similar expressions for $\mathrm{S}'_{\mathrm{F}1}$ and $\Delta'_{\mathrm{F}1}$, it is not necessary to consider explicitly the inclusion of square parts; because the only irreducible self-energy parts are loops with two vertices, (II. Figure 5); and square parts can only be included as sub-parts of included vertex parts. They will thus be allowed for by the substitution of Γ_1 for γ_5. The development of $\mathrm{S}'_{\mathrm{F}1}$ and $\Delta'_{\mathrm{F}1}$ will thus be just the same as in electrodynamics.

Since the infinite constant A″ can be developed as a power series in f, the constant $\delta\lambda$ can be chosen to satisfy

$$\mathrm{A}''(f)+i\delta\lambda(f)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

(This is consistent with the choice of the first term of $\delta\lambda$ made above.) Then no infinities will arise from square parts provided the sum is taken over all such parts at each stage: K^* is equal to K^*_c and K_1 is identical with K. But such summations always are taken in the procedure laid down by Dyson. Thus, if this choice of $\delta\lambda$ is made, no infinities will arise from square parts if Dyson's procedure for electrodynamics is followed for this interaction. In particular Dyson's proof (II. § 7), that the dropping of the other infinite constants is equivalent to a ' charge ' renormalization,

$$f_1 = Z_1^{-1}Z_2Z_3^{\frac{1}{2}}f \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

can be taken over directly, with the trivial addition that $\delta\kappa^2$ must be chosen to satisfy

$$\delta\kappa^2 = (-1/\pi i)Z_3^{-1}A'(f_1). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

The renormalized ' charge ' f_1 can be equated to the finite experimental value and a finite S-matrix is thus obtained for effects arising from this interaction. Strictly speaking the elements of the S-matrix have been derived as a power series in f with finite terms. The convergence of the series itself has not been discussed.

Exactly the same argument applies to the scalar interaction of scalar mesons with nucleons if γ_5 in the above is replaced by $-i\delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is a Kronecker δ-symbol operating on the spinor suffixes of ψ and $\bar{\psi}$.

§ 5. OTHER MESON-NUCLEON INTERACTIONS.

In all the remaining meson-nucleon interactions the condition of § 3 is not satisfied, and the factor corresponding to a meson line in an integral M will contain a double derivative of a Δ_F-function. Thus, for considerations of convergence, a meson line will contribute to M a factor

$$(\Sigma a_i p_\mu^i)(\Sigma a_i p_\nu^i)/((\Sigma a_i p^i)^2+\kappa^2).$$

The new convergence function K^D is defined

$$K^D = F_f - 4(F-n+1). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

This function has the same properties with regard to the convergence of the integrals now under consideration as K had for the interactions previously considered. The expressions (23) and (24) for F_f and F are still valid so that

$$K^D = (3/2)E_f + 2E_b - n + n_b + n_f - 4. \quad . \quad . \quad . \quad . \quad . \quad (43)$$

A primitive graph certainly converges if

$$K^D \geqslant 1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

but primitive graphs will be primitive derivatives if, by (43),

$$(3/2)E_f + 2E_b < n+5. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

Owing to the appearance of n on the right-hand side of this inequality, there is no restriction on the number and degree of the primitive divergents. An irreducible divergent graph may be found for any real process provided sufficient virtual processes are included. The integrals will only converge if the coefficients of all the divergent powers of the basic energy-momenta vanish identically. That this is not always so can be proved by any particular example. Thus for these interactions the renormalization of mass and charge is quite inadequate to remove all the divergences.

The only exception to this is in the vector interaction of neutral vector mesons, in which the derivatives in the factor (17) come from the commutation relations themselves (16) and not from derivatives of the commutation relations as in (15). If the Stückelberg (1938) form is used, in which two subsidiary fields are introduced, it can be seen that the derivatives arise from a vector interaction of the subsidiary scalar field. This can be eliminated by a transformation, and Dyson's treatment of electrodynamics can then be applied without the necessity of introducing a $\delta\lambda$ term. This interaction has already been dealt with elsewhere (Matthews 1949 b), where it is shown that (17) can be replaced by

$$\langle \mathrm{P}(\phi_\mu(x), \phi_\nu(y))\rangle_0 = \tfrac{1}{2}\hbar c\delta_{\mu\nu}\Delta_{\mathrm{F}}(x-y). \quad . \quad . \quad . \qquad (46)$$

§ 6. COMPARISON WITH RADIATION DAMPING.

It is interesting to see how the renormalizing procedure when applied to mesons compares with the much more drastic subtraction procedure of Heitler's (1941) theory of radiation damping *. Heitler's theory has been most successful when applied to the scattering and creation of mesons by nucleons. In the cases where ordinary perturbation theory gives a cross-section which increases with primary energy, the inclusion of Heitler's damping term gives a cross-section which decreases with primary energy. However, this situation does not arise in the scalar or speudo-scalar meson scalar interactions, nor for the vector interaction of neutral vector mesons. (If it did occur, it would be only for longitudinal mesons, and Feynman (1949 b) has proved that the cross-section is zero in this case. The proof is given for photons and is applicable to neutral vector mesons. However, it depends on the cancellation of certain terms, some of which do not occur if the mesons are charged. But for charged vector mesons Dyson's procedure cannot be effectively applied.) Thus the renormalizing procedure is inadequate in all cases where an interesting comparison might have been made, and Dyson's theory, as here applied to meson-nucleon interaction, is not, as might have been hoped, in any important case an effective and preferable alternative to Heitler's.

ACKNOWLEDGMENTS.

In conclusion the author would like to thank Dr. N. Kemmer for his interest and criticism and Mr. F. J. Dyson for a very helpful discussion.

REFERENCES.

CASE, K. M., 1949, *Phys. Rev.*, **75**, 1440.
DYSON, F. J., 1949 a, *Phys. Rev.*, **75**, 486; 1949 b, *Phys. Rev.*, **75**, 1736.
FEYNMAN, R. P., 1949 a, *Phys. Rev.*, **76**, 749; 1949 b, *Phys. Rev.*, **76**, 769.
HEITLER, W., 1941, *Proc. Camb. Phil. Soc.*, **37**, 291.
MATTHEWS, P. T., 1949 a, *Phys. Rev.*, **76**, 684; 1949 b, *Phys. Rev.*, **76**, 1254; 1949 c, *Phys. Rev.*, **76**, 1657.
SCHWINGER, J., 1948, *Phys. Rev.*, **74**, 1492.
STÜCKELBERG, E. C. G., 1938, *Helv. Phys. Acta*, **11**, 225.

---

* This theory has been developed in numerous subsequent papers.

XVII. CORRESPONDENCE

Observation of Unimolecular Growth Steps on Crystal Surfaces.

By L. J. Griffin,
Department of Physics, Royal Holloway College *.

[Received December 16, 1949.]

[Plates VII & VIII.]

An extensive line structure on prism faces of beryl crystals has been revealed by microscopic observation. Evidence is given below to show that these lines represent steps only one unit cell in height; if this is the case the height of the step should be 9·17 Å × cos 30 = 7·9 Å).

The appearance and behaviour of these lines are the same whether the face is examined in its untouched, natural state, or after the deposition of a highly reflecting thin film of silver on the face. The sole difference between the two cases is that the visibility of the lines is much higher when the surface has been silvered. These lines must therefore correspond to genuine geometric surface details.

As nearly as one can tell, these lines follow contour lines of the surface topography. The lines are widely spaced where the surface is relatively flat, and crowd together where there is a steep gradient. The only evident departure from contouring is that occasionally a single line terminates at a point in the face, instead of forming a closed loop. These lines must therefore represent steps on the surface. If the steps are taken to be one unit cell high, the point of termination must be interpreted as the end of a screw dislocation.

An exact survey of the surface topography by interferometric methods should permit of a direct measurement of the step height at each line. So far attempts to apply this method have only sufficed to show that the step height is less than 34 Å. From the fact that no visible kink appears where an interference fringe crosses one of the lines, one may also infer that the step height is less than 20 Å.

Further evidence will be presented below, in the discussion of fig. 3, that the lines represent steps all of the same height. The evidence thus points towards these steps being one unit cell high.

The observation under the microscope of such small steps was surprising, and so contrary to normal practice or prediction, that it was felt that further evidence was needed before the results could be reported. They were not, therefore, made public outside this Department. During

---

* Communicated by the Author.

the Bristol University Summer School on the Theory of Crystal Growth in September 1949, the theory of the behaviour of the steps originating from screw dislocations was developed much more fully by Frank, and it was seen that the correspondence was so exact between the predictions of this theory, developed for unimolecular steps, and the observations on beryl crystals, that one was led to the unavoidable conclusion that the observed steps were, in fact, only one unit cell high.

The technique used in observing these steps consisted in first coating the face to be examined with a film of silver of reflectivity some 70–90 per cent, the silver being deposited under conditions normally used for multiple-beam interferometry (thermal evaporation in a vacuum of pressure about 10^{-5} mm. Hg.). The surface was then examined in reflexion using a normal type of metallurgical microscope, bright field illumination being used with the field iris stopped down to give a fairly narrow illuminating beam, the specimen being racked slightly off focus.

By going slightly off focus when observing the surface, a phase difference is introduced between the diffracted wave fronts, and a certain intensity distribution is produced defining the position of the step edges. The precise nature of the intensity distribution is, of course, dependent on the position of the focal plane of the objective, non-complementary intensity distributions being produced on opposite sides of the true focal plane. In general, it has been found best to work with the edge imaged by an intensity distribution consisting of a broad minimum between two narrow maxima (see figures). These unimolecular step edges have been observed with objectives of N.A. ranging between 0·058 and 1·3, the visibility of the effect apparently decreasing with decreasing numerical aperture of the objective : it is, however, difficult to determine whether this decrease in visibility is primarily due to the decrease in the observing numerical aperture, considered as the collector of a cone of light of certain angle, or is caused, partly or entirely, by other factors dependent on the value of the numerical aperture.

The deposition of a reflecting silver layer, some 250–400 A.U. thick, greatly increases the visibility of the diffraction effect produced by the step edges. It does not seem to alter the nature of the intensity distribution in the image of the edges as obtained from an unsilvered face. This was to be expected from the evidence obtained from multiple-beam interferometry (Tolansky 1948) that such silver films, when deposited under correct conditions, contour the surface of the substrate to within very close limits.

It would seem probable that improvements could be made in the technique of observation which would further increase the sensitivity.

The topography of prism faces of beryl crystals is in general complex, with a tendency to form occasional low plateaux of a typical " razor-blade " shape, similar in outline to the outer loop of fig. 2. On the surface of two such main plateaux the two very simple cases shown in figs. 1 and 2 were found.

Fig. 1 (magnification 625 ×) shows the simplest case of a single screw dislocation, imaged here as a dot, from which runs a spiral edge, as predicted by the theory (Frank 1949 a). The *c*-axis of the crystal is here, as in all the figures, parallel to the almost straight edges. The light area is a region of very coarse structure, the general, irregular, and randomly distributed, surface markings being probably due to the deposition of foreign material after the beryl crystal had ceased to grow. This irregular and random marking is of general occurrence on faces of beryl, varying considerably in density.

Fig. 2 (magnification 1050 ×) shows the next most simple case, that of growth from the step joining a right and a left handed screw dislocation (Frank 1949 b). According to the theory unclosed loops like the inner one, attached to a pair of dislocations, extend during growth of the crystal in such a way as to generate successive closed loops, exemplified by the outer one. (The triangular area cutting into the closed loop probably consists of an over growth of foreign crystalline material). The two dislocations in this case are also imaged as dots, this being the usual type of imaging on these prism faces.

Figs. 3 and 4 show the more normal type of behaviour seen on beryl, the above simple cases being exceptional. In fig. 3 (magnification 625 ×) the steps advancing from the right are split into three sections by a succession of three crystal inclusions, which act as obstacles. Of the three sections, the central and lower sections first combine again to form a single section. Counting from the last edge before division commences on the right, there are 37 edges in both the central and lower sections before junction occurs again. Similarly, considering the junction beyond the third obstacle, there are 57 edges in both the upper and lower sections between the first division of the edges and the first complete edge. It must, therefore, be that, in general, each step is a like unit step. For a step one unit cell high this type of behaviour is immediately understandable, and in fact necessary, but for steps of two or more unit cells it would appear very probable that in these circumstances the steps would either split into smaller units or combine to form multiple steps.

The slow closing in of the edges beyond any inclusion, and the rapid elimination of the re-entrant angle formed at the junctions of the different sections, is typical.

Fig. 4 (magnification 500 ×) shows the effect produced by a distribution of disclocations "dominated" by the steps originating from another group, far off the picture to the left. Some of its features are discussed by Frank, below.

Similar behaviour to that reported above has been seen on two other crystals of beryl, in these cases of typical emerald appearance and habit, while the above crystal was of aquamarine type. The three crystals came from widely different localities (*e. g.* the Urals and California), so that this would seem to be a general type of behaviour of beryl crystals.

Fig. 1.

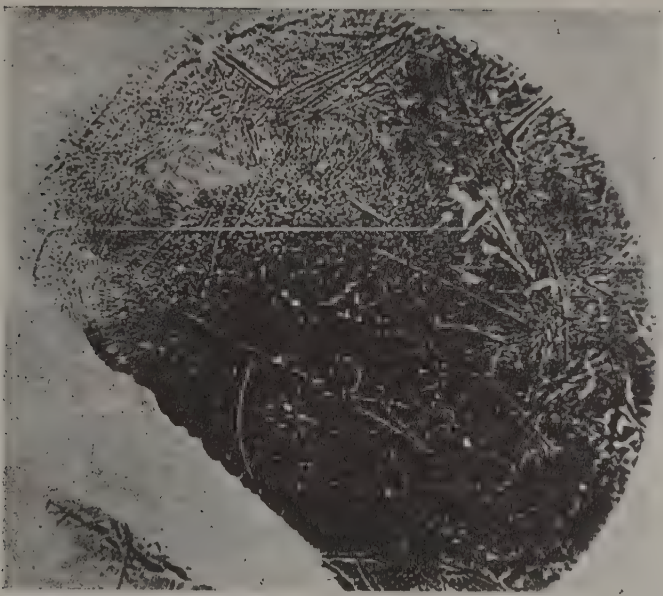
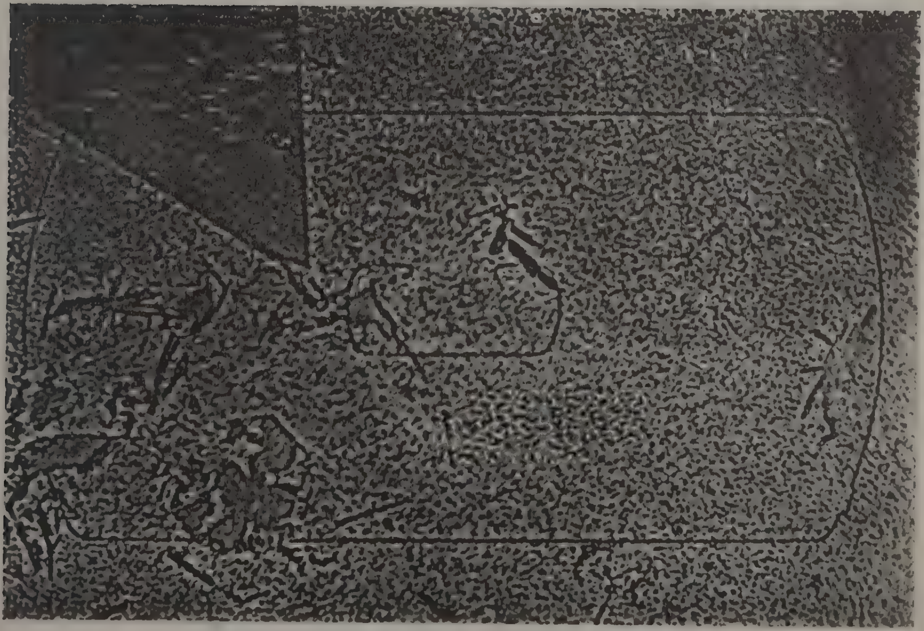


Fig. 2.



Figs. 1, 3 and 4 negative. Fig. 2 positive.

Fig. 3.

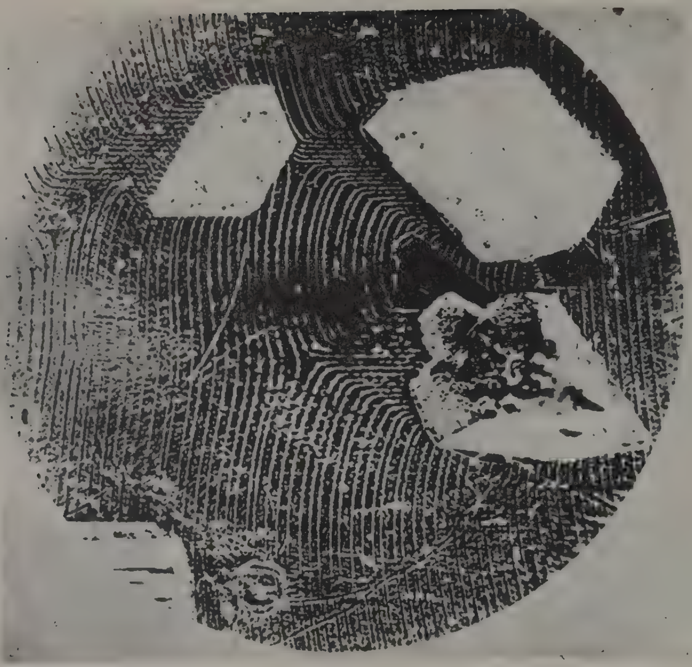
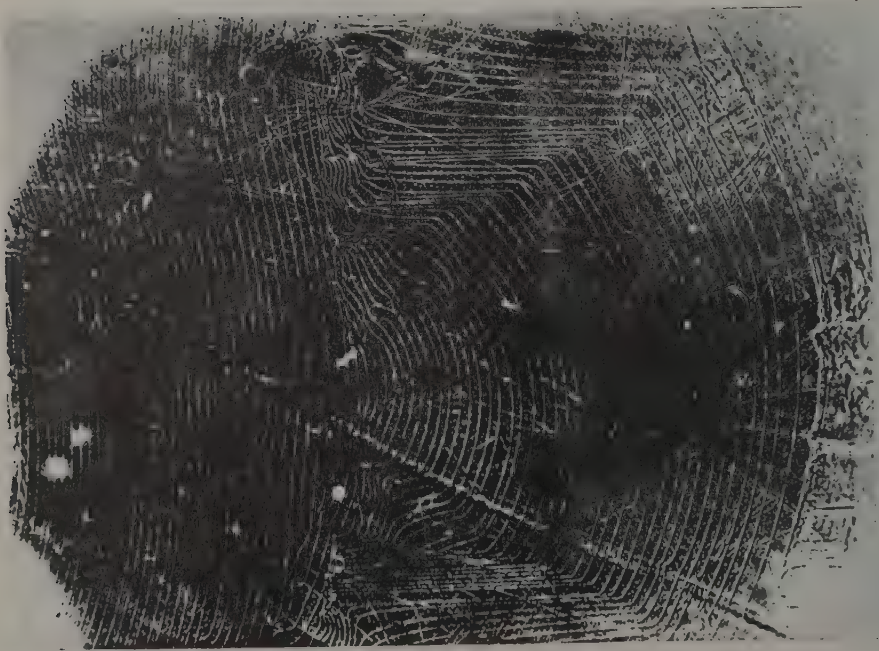


Fig. 4.



Taking into account the shortcomings of two-dimensional nucleation theory (Burton, Cabrera and Frank 1949, Cabrera and Burton 1949), the general validity of the dislocation theory of crystal growth would seem to be established. Further confirmation of the theory should be obtainable not only by observations of similar static phenomena on other types of crystals, but also by observing successive small amounts of growth or solution on faces of crystals, the topography of which had been previously determined. By this means a kind of cinematograph picture could be built up of the behaviour of the steps. In the case of beryl, further growth is not possible, but it is hoped to observe solution effects and thereby test the theory further.

Some small rod-like structures can be seen in the lower part of fig. 4. This rod type of structure is of common occurrence on the beryl crystals examined and occurs invariably on steep gradients running in the general direction of the *c*-axis. The rods are often found with one or more dislocations at their base and usually show a somewhat irregular step structure.

The difference in density of occurrence of screw dislocations per unit area can be considerable, both between different areas of the same face, and crystallographically equivalent faces of the same crystal. Thus there are some 100 screw dislocations in an area of $7{\cdot}6 \times 10^{-3}$ sq. mm. of one face, on other parts of which areas of several square millimetres show no sign of dislocations. An equivalent prism face of the same crystal shows no sign of dislocations over an area of several square centimetres. There is a definite tendency for dislocations to be closely spaced along lines either roughly parallel or perpendicular to the *c*-axis, as one would expect from general dislocation theory.

ACKNOWLEDGMENTS.

I am grateful to Professor S. Tolansky for his interest and encouragement during the course of this work. My thanks are due to the Mineralogy Department, British Museum, for the loan of several specimens ; also to the Department of Scientific and Industrial Research for the provision of a maintenance grant.

REFERENCES.

BURTON, CABRERA and **FRANK**, 1949, *Nature*, **163**, 398.
CABRERA and **BURTON**, 1949, *Discussions of the Faraday Society*, No. 5, 40.
FRANK, F. C., 1949 a, *Discussions of the Faraday Society*, No. 5, 48.
FRANK, F. C., 1949 b, *Ibid*, 67.
TOLANSKY, S, 1948, *Multiple-Beam Interferometry of Surfaces and Films* (O.U.P.), 63, 166.

Note on the Structure of a Crystal Surface.

By F. C. Frank,
H. H. Wills Physical Laboratory, Bristol*.

[Received December 23, 1949.]

The observations reported by Griffin (1950) provide an unexpected direct confirmation of the theoretical conclusion that dislocations play an essential role in the growth of crystals. Indeed, they provide the first direct visual evidence of the existence of dislocations themselves. They also offer promise of a much closer insight into the actual conditions under which the growth of natural crystals occurred. If it also turns out that a similar technique will reveal details of "molecular" thickness on other crystals, growing in the laboratory, the possibilities for enlarging our knowledge about crystal growth processes are, of course, enormous.

The quantitative theory of crystal growth which has been developed in various papers by Burton, Cabrera and the writer (1949, 1950) (some of which, as yet unpublished, were presented at the Bristol Summer School on the Theory of Crystal Growth) has so far dealt mainly with crystals formed from spherical or cubic molecules, with growth occurring on crystal faces of high symmetry. In qualitative principle the theory applies equally, though with complications, to the $(10\bar{1}0)$ face of Beryl, studied by Griffin. There appears to be no reason to doubt that the essential building unit of Beryl is the $Si_6\ 0_{18}$ ring with its attendant two aluminium and three beryllium ions, making half the unit cell found by Bragg and West (1926). These units may be regarded as "pill-boxes", 7·9 Å. in diameter and 4·6 Å. in height, close-packed side by side in the basal planes of the crystal, and one on top of the other in the direction of the hexagonal axis. In this direction alternate $Si_6\ 0_{18}$ rings differ in orientation by 25°. This alternation secures the best coordination of oxygens about the cations, and doubles the X-ray unit cell, but may be considered a secondary detail of the building principles so far as we are concerned. On a $(10\bar{1}0)$ face these "pill-boxes" lie on their sides packed in flat layers in rectangular array, and growth would be expected to proceed by the spreading of such "monomolecular" layers by accretion of new units at their edges. At the low supersaturations essential for the growth of good crystals such layers would not develop on a free flat surface except from the points of emergence of dislocations with a component of the displacement vector normal to the surface ("screw dislocations" for short, though it is not essential that they should be pure screw dislocations). The way in which the edges of monolayers develop as "growth spirals" centred on the dislocations have been described elsewhere (Frank 1949), and beautifully illustrated in Griffin's figs. 1 and 2.

---

[*] Communicated by the Author.

The dependence of the two-dimensional crystal growth rate on the orientation of the step line imposes a characteristic distortion of the growth spirals and loops, exhibiting the symmetry of the crystal face. A fairly simple theory can be given to account for the shape assumed by each growth layer as it expands, the characteristic " razor-blade " shape seen on the beryl crystal being a typical example. In essence it is the same theory as was developed earlier (see for example, Friedel (1926)) for three dimensional crystal growth on the assumption that the growth rate at any point on a crystal surface depends only on the local crystallographic orientation of the surface, and its environment. We now appreciate that this is a far from correct assumption in that case, but in two dimensions it should be much more nearly true. In the latter case, a straight edge corresponds to a sharp minimum in the growth rate as a function of orientation. Truly sharp minima, and therefore truly straight boundaries are not in general to be expected. When the rate of advance, v, of the step, depends on its orientation ψ according to the law

$$v=v_0(1+B\psi^2),$$

we can show that the resulting boundary will have a radius of curvature $(1+2B)$ times its distance from the centre of the figure. A simple theory, applicable in certain circumstances which are particularly likely to occur in growth from solution, makes the rate of advance v proportional to the number of " Frenkel kinks " in the step line. Then it follows that $B=e^{\varphi/kT}/8$, where ϕ is a neighbour-neighbour interaction energy which we may take to be $\frac{1}{3}$ of the latent heat of solution. In this case the nearly straight edge of the razor-blade shape in Griffin's figure (2) has an apparent radius of curvature 100 times its distance from the centre, though it may be more (*i.e.* the apparent curvature is possibly due to optical distortion). Thus ϕ/kT is 6, or more. By contrast, the growth figures on a synthetic emerald grown hydrothermally by Nacken (1948) (Photographs by Griffin, not reproduced here) are similar, but have strongly curved edges to the " razor-blade " indicating $B=8$, whence $\phi/kT=4{\cdot}16$. Supposing the natural and synthetic crystals grew from similar media, we may deduce that the absolute temperature of growth of the natural crystal was only about two-thirds Nacken's working temperature. Information of this is unfortunately lacking, but the indication that the natural crystal probably grew at a moderate temperature may be of interest.

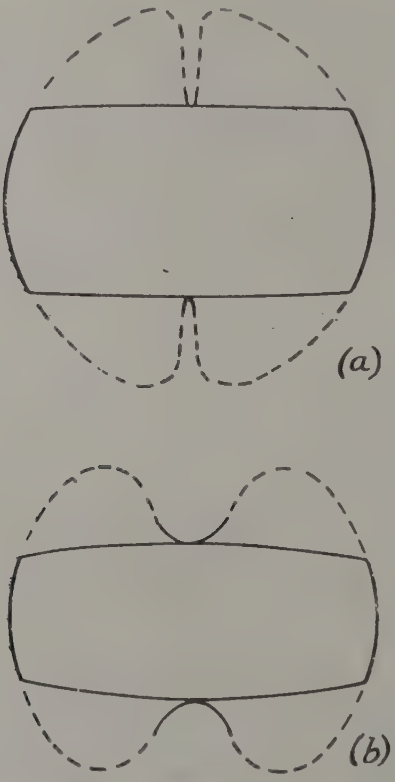
The two ends of the razor-blade are very nearly arcs of the same circle both for natural beryl and Nacken's crystals, indicating that the growth rate is practically independent of orientation for the range of orientations occurring in them (±30° from the normal to the hexagonal axis in the case of the natural crystal).

Fig. 5* shows polar diagrams of two dimensional growth rate which would account for the observed two-dimensional growth figures. These polar diagrams are naturally uncertain for the range of orientations which do not appear in the figures : in such ranges one can only say for certain that the polar diagram lies outside the circles whose diameters are the lines connecting the centre to the corners of the figure.

---

* Figures in this letter are numbered from 5 onwards to simplify reference to figures in Griffin's article, figs. 1 to 4.

Fig. 5.

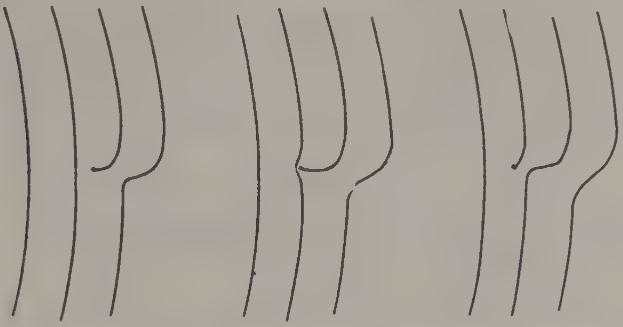


The "razor-blade" outline of a growth layer on the (10$\bar{1}$0) face of beryl (hexagonal axis horizontal) and the associated polar diagram of rate of growth as a function of boundary orientation, (*a*) for the natural crystal, (*b*) for Nacken's synthetic crystal. The form of the polar diagram is uncertain where represented by a dashed line and coincides with the "razor-blade" outline on circular arcs at the ends.

Observation of the spacing between turns of a growth spiral, or of its form where it joins the dislocation, enables us to measure a characteristic length, "the size of the critical nucleus", whose order of magnitude is $a\phi/kTln\alpha$ where a is a lattice spacing of the crystal and α is the saturation ratio. From fig. 2 this length is found to be about 1 micron, whence $\phi/kTln\alpha$ is about 10^3. In conjunction with the former estimate of ϕ/kT we may deduce that the degree of supersaturation at which growth took place was about 1 per cent. These estimates are not intended to be final or accurate, but are meant to indicate the kind of information which we should be able to derive from the type of observation Griffin has been able to make.

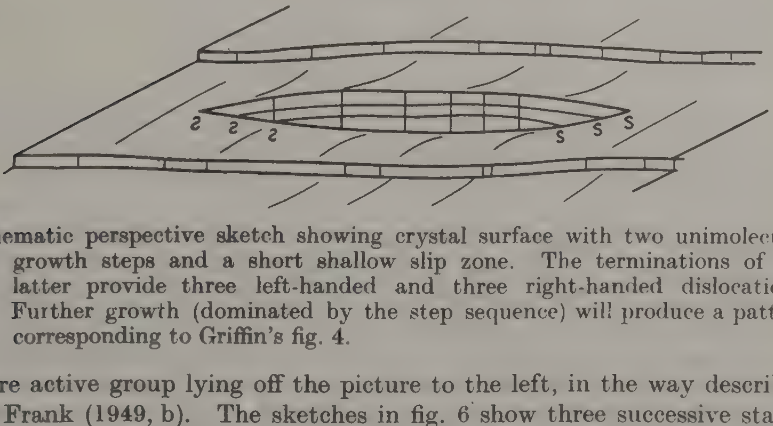
Whereas Griffin's figs. (1) and (2) show the simplest kind of growth initiation centres, he has other photographs showing initiation by larger groups of dislocations, such as occur in his fig. 4, though here they are not acting as an initiating group because they are " dominated " by a more active group lying off the picture to the left, in the way described by Frank (1949, b). The sketches in fig. 6 show three successive stages in the passage of a dominating sequence of growth fronts across a single dislocation, stages which may be recognized at various places in fig. 4. Fig. 4 also shows, as well as the randomly distributed dislocations, two groups lying approximately on a straight line, making a closely spaced row of right-handed dislocations at the top of the picture, and of left handed dislocations near the bottom. This is the kind of distribution of screw

Fig. 6.



Three successive stages in the passage of a dominating sequence of growth steps over a dislocation. Each line represents a unimolecular step, down to the right. The next stage in this succession will resemble the first.

Fig. 7.

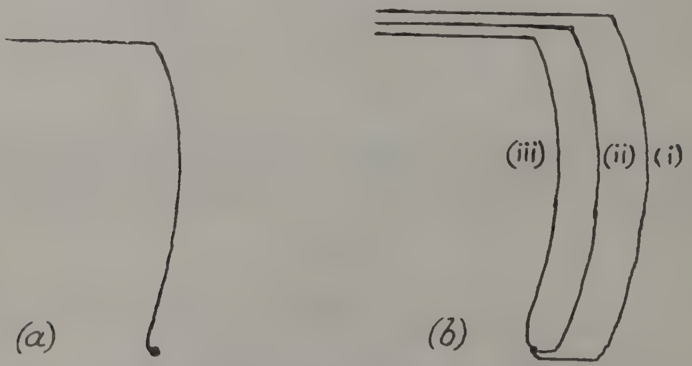


Schematic perspective sketch showing crystal surface with two unimolecular growth steps and a short shallow slip zone. The terminations of the latter provide three left-handed and three right-handed dislocations. Further growth (dominated by the step sequence) will produce a pattern corresponding to Griffin's fig. 4.

dislocations which results from the formation of a limited slip-zone, as illustrated in fig. 7. One may readily see how subsequent growth on a surface of this configuration, under the domination of a close sequence of step-lines originating elsewhere, produces the general pattern of fig. 4.

It is interesting to consider the origin of such slip zones, and similar arrangements of dislocations to be found elsewhere in pictures taken by Griffin, in which, however, the line is parallel to the hexagonal axis, instead of transverse to it. A likely cause is the stress resulting from inhomogeneity of ionic replacement in the crystal. The fact that the crystal is coloured is sufficient evidence that ionic replacement occurs. A change in the composition of the solution will produce inhomogeneity in the crystal, which must evidently lead to lattice curvature. Now, consideration of the growth of a curved crystal from a mixed solution shows that it will suffer oscillations in the lattice spacing, with a wave length of the order of magnitude of its shortest dimension, accompanied by considerable stresses. Dislocations already present may move under these stresses, but, additionally, the surface stress in new growing layers (particularly shear stress and tension) can, relatively easily, produce surface faults, which become true dislocations when buried under additional layers. This is probably the chief source of new dislocations, lineage boundaries and the like, in slowly growing crystals. Of course, the initial dislocations needed to make the first crystal seed capable of growth are generated with ease at the high supersaturation needed for nucleation.

Fig. 8.



(*a*) A form of growth front occasionally observed ; (*b*) the normal form of growth spiral (i) is transformed into the shape (iii), with reversed curvature, as in 8(*a*), by a slight dissolution following normal growth.

On certain crystals, Griffin finds that the growth fronts have a reversed curvature near their terminations on dislocations, as in fig. 8(*a*). This appears to indicate that the last phase in the development of the crystal was a slight dissolution, before all change ceased. Fig. 8(*b*) shows how the normal growth spiral (i), retracting in dissolution by way of stage (ii), proceeds to wind the spiral the other way, close to the dislocation, as in stage (iii) which corresponds to 8 (*a*).

H. H. Wills Physical Laboratory,
The University, Bristol.
F. C. Frank.

References.

Bragg and West, 1926, *Proc. Roy. Soc.*, A **111**, 691.
Burton, Cabrera, and Frank, 1949, *Nature*, **163**, 398.
Burton and Cabrera, 1949, *Crystal Growth* (Discussions of the Faraday Society No. 5), 33.
Burton, Cabrera, and Frank, 1950 to be submitted to the Proceedings of the Royal Society.
Cabrera and Burton, 1949, *Crystal Growth*, 40.
Frank, 1949, *Crystal Growth* (*a*) 48, (*b*) 67.
Friedel, 1926, *Lecons de Cristallographie*, 262, Paris, Berger-Levrault.
Griffin. This Journal, p.
Nacken, *see* Van Praagh, 1948, *Research I*, 458.

Transfer Rates for Surface Films of HeII.

By E. J. Burge and L. C. Jackson,
H. H. Wills Physical Laboratory, University of Bristol*.

Published measurements of the rate of transfer of HeII by the surface film fall into two groups. On the one hand there are the pioneer studies of Daunt and Mendelssohn (1939) which showed that the rate depends only on the smallest periphery above the liquid helium level and on the temperature, increasing from zero at the λ-point to a flat maximum of about $7{\cdot}5\times10^{-5}$ cm^3/sec/cm. at 1·5° K. These observations were substantially confirmed by the work of Webber, Fairbank and Lane (1949) and the later work of Atkins (1949).

On the other hand Atkins in his earlier work (Atkins 1948 a) observed rates up to ten times those of Daunt and Mendelssohn and showing a marked dependence on the distance of the higher helium level from the rim of the beakers but independent of the distance between the two liquid levels. He used a capillary beaker terminating in a spherical bulb and the results were attributed to temperature gradients and possibly the geometry of the vessel.

Similarly de Haas and van den Berg (1949) have recorded high creep rates and a dependence of the rate on the difference between the helium levels. They found the rates of *emptying* the beaker were those to be expected for a syphoning process, *i.e.* dependent on the difference of height of the helium levels This type of process will not, however, explain the rates of *filling* initially observed by Atkins, and confirmed later at Oxford and at Bristol (for contaminated surfaces), which were independent of the difference of height of the levels but dependent on the height of

---

* Communicated by the Authors.

the rim of the beaker above the highest helium level. Their experiments were carried out in the absence of radiation and they attribute the difference between their results and those of Daunt and Mendelssohn to this factor.

Bowers and Mendelssohn (1949) showed that absence of radiation had no effect on the normal creep rates, and that the high creep rates observed by Atkins in his first experiments were due to films of solidified gases (*e. g.* air) on the walls of the vessels. In his later experiments Atkins has confirmed the correctness of this explanation. It may be of interest to record our own observations on creep rates collected during the course of our experiments on helium II films (Jackson & Burge, 1949) and designed to test the similarity of our experimental conditions to those of Daunt and Mendelssohn in their measurements of creep rates and film thickness. These observations were made with a cylindrical beaker enclosed in the combined uniform temperature enclosure and radiation shield shown in fig. 1 (*a*). The walls are silvered except for two narrow longitudinal slits. A half-watt neon lamp with a filter of Chance's ON19 glass served as the source of illumination. The creep rates agreed closely with those of Daunt and Mendelssohn.

Fig. 1 (*a*)

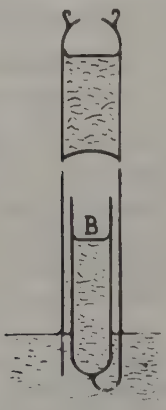
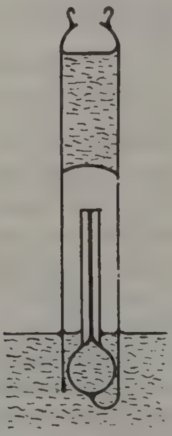
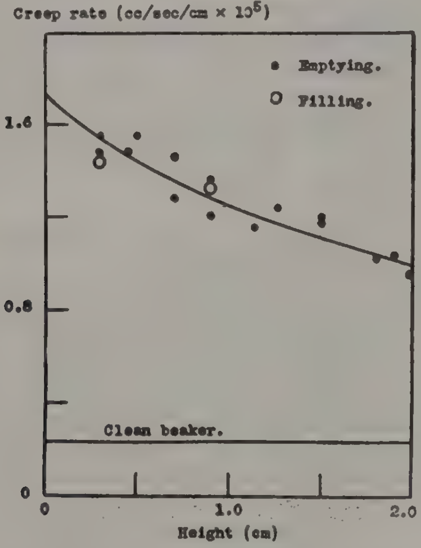


Fig. 1 (*b*)



After the publication of Atkins' first note (Atkins, 1948 a) the cylindrical beaker of fig. 1 (*a*) was replaced by one consisting of a glass capillary tube, length 2·2 cm., internal diameter, 0·04 cm. terminating in a spherical bulb of diameter 0·8 cm. fig. 1 (*b*). The creep rates observed again agreed very well with those of Daunt and Mendelssohn. The appearance of the note of Bowers and Mendelssohn led us to repeat our observations (1) with the "clean" Atkins type beaker, and (2) after admitting some 50 cm^3 of air to the cryostat, a small portion of which condensed upon the beaker. The creep rate with the "clean" beaker was again normal but the maximum rate with the contaminated beaker was some seven times greater and a marked variation with height was observed (fig. 2; $T=1{\cdot}90^\circ$ K). We confirmed Atkins' observations that the filling rates depended on the height of the rim above the highest helium level and not on the difference of the levels. In some of the experiments with the contaminated beaker the observations were continued until the levels of the liquid helium inside and outside the beaker were the same. The inner level was then found to overshoot and the oscillations first described by Allen and Misener (1939) and studied in detail by Atkins (1948 b) were observed. In one experiment twelve complete periods were timed, giving an average value for the period of oscillation of 10·4 secs.

Fig. 2.



Assuming the film thickness to be inversely proportional to the height and inserting the above value of the period of oscillation in the formula given by Atkins, the film thickness at a height of 1 cm. was calculated. The value obtained, $d=11{\cdot}16 \times 10^{-6}$ cm. ($T=1{\cdot}07^\circ$ K) agrees well with the early results of Atkins. Unfortunately similar observations had not been made with the clean beaker.

REFERENCES.

ALLEN, J. F., and MISENER, A. D., 1939, *Proc. Roy. Soc.* A. **172**, 467.
ATKINS, K. R., 1948 a, *Nature*, **161**, 925.
ATKINS, K. R., 1948 b, *Thesis*, Cambridge.
ATKINS, K. R., 1949, *Rep. Int. Conf. Phys. Very Low Temps.*, Camb. Mass.
BOWERS, R., and MENDELSSOHN, K., 1949, *Nature*, **163**, 870.
DAUNT, J. G., and MENDELSSOHN, K., 1939, *Proc. Roy. Soc.* A. **170**, 423.
DE HAAS, W. J., and VAN DEN BERG, G. J., 1949, *Rev. Mod. Phys.*, **21**, 524.
JACKSON, L. C., and BURGE, E. J., 1949, *Nature*, **164**, 660.
WEBBER, R. T., FAIRBANK, H. A., and LANE, C. T., 1949, *Phys. Rev.*, **76**, 609.

XVIII. *Notices of New Books and Periodicals received.*

Introduction to Atomic Physics. By S. Tolansky. Third Edition. [Pp. 371.] (Longmans Green.) Price 15*s.*

The appearance of this text for the degree student is welcome because the second edition with its 1946 appendix was already somewhat out of date.

New sections on recent developments have been included in various subjects. The chapters on cosmic rays and the cyclotron type of machine have been brought more up to date and that on the nucleus has been enlarged. Of the two new chapters added, one deals entirely with the neutron and its properties; the second concerns nuclear fission and atomic energy.

It is in one way unfortunate for the author that some of the subjects dealt with are developing at so remarkable a rate that the information is already out of date, e.g., in the sections dealing with mesons and the photographic study of nuclear and electron tracks.

In our experience students have been almost unanimous in praising the second edition as a stimulating source of valuable information. They will continue to do so with the new edition. The work is relatively free from the faults that marred the first edition, though the reviewer notices one unfortunate statement on page 318 in the paragraph on the betatron, namely, "The mass increase limits the ultimate velocity". It is feared that the student will take this to mean that relativistic change of mass sets a limit to the energy obtainable, which, of course, is untrue. H. H. P.

Ions, Electrons and Ionizing Radiations. By J. A. Crowther. (8th Edition.) [Pp. 322.] (Edward Arnold.) Price 21*s.* 0*d.*

That this work has been accepted for many years as a standard text-book is shown by the fact that the demand for it has required an eighth edition. One reason is that it is eminently readable and contains no information which is superficial or unreliable. The main changes which have been introduced since the last edition, eleven years ago, have naturally been in the field of nuclear physics and atomic energy. In this field there is such a wide choice of material that there is room for individual preference. The reviewer would like to have seen some reference to Powell's development of the photographic method for the quantitative study of nuclear particles. It is now ten years since he pointed out the potentialities of this method, which are now being so effectively exploited as to speed up the tempo of advance many times. On the other hand, one can welcome the author's refusal to throw on the scrap-heap descriptions of early discoveries in the vain hope of circumscribing completely the subject of atomic physics within the limits of a single volume. This point of view will be shared by many teachers, particularly those who, like the reviewer, have watched the subject develop through the past fifty years.

A. M. T.

Faraday's Discovery of Electromagnetic Induction. By T. Martin. [Pp. 160. (London: Arnold.) Price 9*s.*

The author of this very readable account of a great discovery and the events which led up to it is the editor of Faraday's diary published before the war. In a connected story, illustrated by excerpts and drawings from the diary, readers can catch some of the atmosphere of scientific investigation a century and more ago and sense the working of mind and intuition of a great experimenter.

The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XIX. *Note on Potentials Derived from Axial Values in Electron Optics.*

By **Feodora Berz**, L. ès Sc.,
Mullard Electronic Research Laboratories †.

[Received October 31, 1949.]

SYNOPSIS.

The determination of electrode systems which produce a precalculated distribution along the optic axis is examined. It is shown that an analytic potential distribution is completely determined when its values along any arbitrarily small interval on the axis are exactly known. On the other hand if, as usual, the value of the potential along a finite stretch of the axis is prescribed, within arbitrarily small but finite limits, this is compatible with an infinity of different potential distributions, which can be constructed without any singularity at finite distances, and which can assume prescribed values at any number of points outside the axis of symmetry.

§ 1. INTRODUCTION.

THE determination of electrode systems to produce a precalculated potential distribution along the optic axis is a standard problem of electron optics. By a well-known theorem, the potential on the axis uniquely determines the distribution outside it. This theorem, though very often quoted, is not always properly understood, especially in its practical implications. The reason is that in this extrapolation problem the difference between the rigorous, mathematical, and the everyday or practical concept of a function, which is often overlooked in physical problems, has far reaching consequences.

In order to make this clear, it will be shown in this paper, first that if the concept of function is used in the *mathematical* sense, the potential function in the plane is determined by its values on a stretch smaller than any prescribed length (its values along the whole axis being either redundant, or leading to non-analytic solutions). Second, it will be shown that if the concept of the "given function" is used in its *practical* meaning, *i. e.* if the potential on the axis is known within finite but arbitrarily small limits, the potential problem admits an infinity of solutions.

The chief practical consequence of these results is that in many electron optical problems, where the exact realization of a precalculated potential distribution on the axis leads to inconvenient or even impossible shapes, very small variations of the data, which do not appreciably influence the electron trajectories, may result in electrode systems far more convenient for practical application.

---

† Communicated by Dr. D. Gabor.

For convenience we shall limit ourselves to plane electrostatic fields, with potential distributions symmetrical with respect to the OX axis. The case of rotational symmetry is more complicated, but the general results obtained can be transferred without alteration, as will be briefly summed up in §4, at the end of this paper.

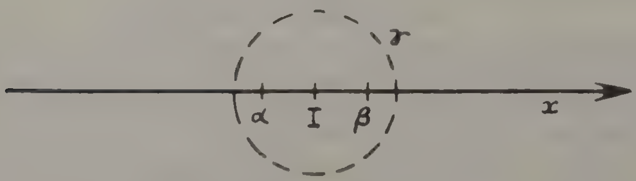
§ 2. RIGOROUS MATHEMATICAL SOLUTION—THEOREM OF UNIQUENESS.

Consider a potential distribution $V(x, y)$ which can be expounded in Taylor series in the neighbourhood of any given point, with the possible exception of some isolated singularities. Such distribution will be called "analytic", by an extension of the theory of functions of a complex variable.

In the absence of space charge, case which will only be considered here, the function $V(x, y)$ must satisfy Laplace's equation

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Fig. 1.



If $v(x)$ is the potential distribution along the OX axis, the solution in the whole plane is

$$V(x, y)=\mathrm{Re}\,v(x+iy), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $\mathrm{Re}\,v(x+iy)$ stands for the real part of the function $v(x+iy)$.

Another well-known formula giving $V(x, y)$ as a power series, and valid in the neighbourhood of the OX axis, is

$$V(x, y)=v(x)-\frac{y^2}{2!}\cdot\frac{d^2v}{dx^2}+\frac{y^4}{4!}\cdot\frac{d^4v}{dx^4}+\ldots+\frac{(-1)^n y^{2n}}{2n!}\cdot\frac{d^{2n}v}{dx^{2n}}+\ldots . \quad (3)$$

Formula (3) can be derived from formula (2), by expanding it in a Taylor series.

Theorem of uniqueness: if two potential functions assume the same value along any arbitrarily small interval of the axis of symmetry, they are identical.

The proof may be directly obtained from the corresponding theorem concerning the analytic functions of a complex variable.

Let $V(x, y)$ and $V^*(x, y)$ be two potential distributions which assume the same value along a small portion $\alpha\beta$ of the OX axis (fig. 1), and let $v(x+iy)$ and $v^*(x+iy)$ be the corresponding complex functions defined above.

$$V(x, 0)=v(x), \qquad V(x, y)=\mathrm{Re}\,v(x+iy),$$
$$V^*(x, 0)=v^*(x), \qquad V^*(x, y)=\mathrm{Re}\,v^*(x+iy).$$

Consider now the analytic function $\phi(x+iy)$ such that

$$\phi(x+iy)=v^*(x+iy)-v(x+iy).$$

Expanding $\phi(x+iy)$ in Taylor's series, with a point I inside $\alpha\beta$ taken as origin, we have

$$\phi(x+iy)=\phi_i+z\frac{d\phi_i}{dz}+\frac{z^2}{2!}\frac{d^2\phi_i}{dz^2}+\ldots+\frac{z^n}{n!}\frac{d^n\phi_i}{dz^n}+\ldots, \quad . \quad . \quad (4)$$

where

$$z=x+iy,$$

and the subscript i refers to values taken at the point I.

Since $\phi(x+iy)$ is an analytic function, the value of its derivatives at any given point does not depend on the direction along which this derivative is taken. Therefore

$$\frac{d\phi_i}{dz}=\frac{\partial\phi_i}{\partial x}, \qquad \frac{d^n\phi_i}{dz^n}=\frac{\partial^n\phi_i}{\partial x^n}.$$

On the other hand $\phi(z)$ is equal to zero along $\alpha\beta$ by definition. Hence its derivatives with respect to x at the point I, and therefore also all the terms of the Taylor series (4), will vanish. Thus $\phi(x+iy)=0$ inside the circle of convergence γ of the series (4), whose centre is I (fig. 1). Using analytic continuation it could similarly be shown that $\phi(x+iy)\equiv0$ everywhere. Hence

$$v^*(x+iy)=v(x+iy)$$

and

$$V^*(x, y)=V(x, y).$$

Thus an analytic potential distribution is completely determined by its values along an *infinitesimal portion* of the axis of symmetry.

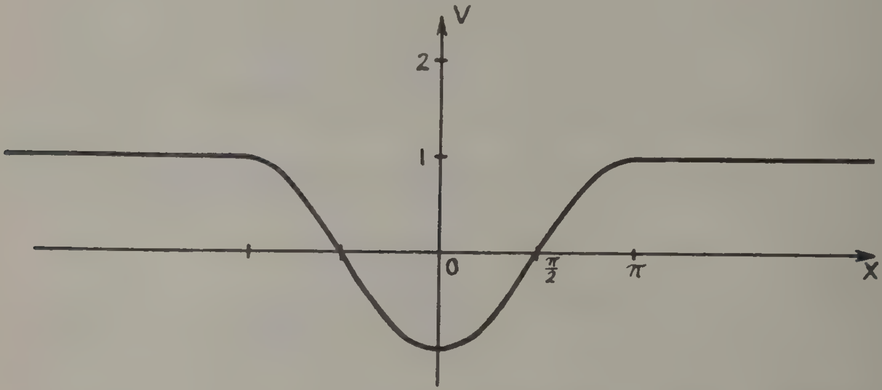
It has to be noted that formulæ (2) (3) tend to give the false impression that the potential distribution along the whole axis OX can be chosen in an arbitrary way. If $v(x+iy)$ is analytic, its values along the whole axis OX are implicitly defined by its values along an infinitesimal portion of this axis. On the other hand if $v(x+iy)$ is not analytic along the axis OX taken as a whole, $V(x, y)$ as defined by (2) or (3) will not be analytic in the neighbourhood of the OX axis.

As an example consider the function $V(x, 0)$ defined by

$$V(x, 0)=-\cos x \quad \text{for} \quad -\pi\leqslant x\leqslant\pi.$$
$$V(x, 0)=1 \quad \text{for} \quad x>\pi \quad \text{and} \quad x<-\pi.$$

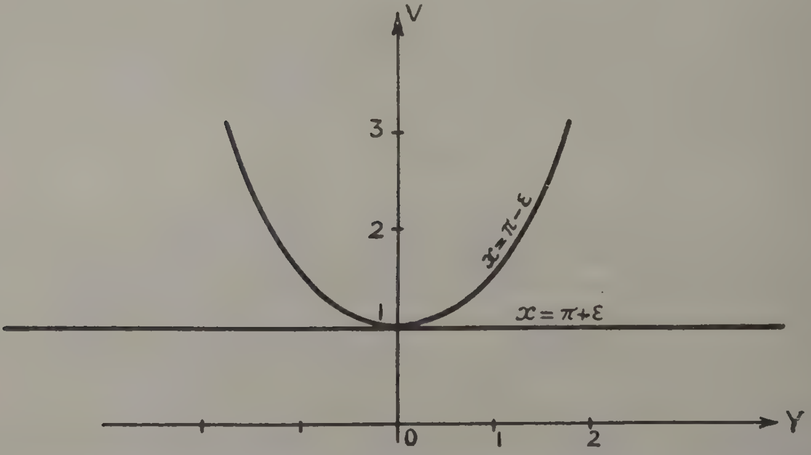
Such a function $V(x, 0)$ is continuous as well as its first derivative (see fig. 2*a*), but it is not analytic as a whole, on account of the discontinuity of higher derivatives at the points $x=\pm\pi$, $y=0$. The potential in the plane, as derived from (2) or (3), is given by

Fig. 2*a*.



Potential variation along the OX axis. $V(x)=-\cos x$ for $-\pi<x<\pi$; $V(x)=1$ for $x>\pi$ and $x<-\pi$.

Fig. 2*b*.



Potential variation along the parallels to the OX axis corresponding to $x=\pi-\epsilon$ and $x=\pi+\epsilon$.

$$\begin{aligned} V(x, y)&=-\cos x \cosh y \quad &&\text{for} \quad -\pi\leqslant x\leqslant\pi, \\ V(x, y)&=1 &&\text{for} \quad x>\pi \quad \text{and} \quad x<-\pi. \end{aligned}$$

Such a potential is discontinuous along the lines parallel to *oy*, for $x=-\pi$ and $x=+\pi$, since for $\epsilon>0$

$$\lim_{\epsilon\to 0} V(\pi-\epsilon, y)=\cosh y,$$
$$\lim_{\epsilon\to 0} V(n+\epsilon, y)=1,$$

and similarly at $x=-\pi$, (fig. 2*b*).

§ 3. APPROXIMATE PRACTICAL SOLUTIONS.

The uniqueness theorem appears to impose very drastic restrictions on the solution of potential problems. It states that potential distributions, even if they are continuous, cannot be arbitrarily prescribed on the axis of symmetry, if an analytic solution is required, and that no alteration of an analytic potential distribution along a finite portion of the axis can be made, without introducing discontinuities, and other very serious singularities.

However these theoretical limitations are of little consequence in practical cases for two reasons. The physicist wants to obtain a given potential distribution along a *finite* portion of the axis of symmetry, and not along the whole axis. Moreover, this potential distribution need be realized only *appoximately*. Physical quantities are never measured exactly, and rigorous coincidence is a meaningless concept in physics.

By a theorem of Weierstrass (Goursat 1927), if a function is continuous over a finite interval of the real axis, it can be approximated over this interval with any accuracy by a polynomial †. The following propositions can be derived from this theorem.

Proposition 1.

Any continuous potential distribution, prescribed along a finite section of the axis of symmetry, can be realized with any degree of approximation by an infinity of different potential distributions.

Proposition 2.

These potential distributions can be constructed without any singularity in the plane.

Proposition 3.

These potential distributions can be constructed with prescribed values at any number of points outside the axis of symmetry.

To prove these propositions consider a real polynomial $P(x)$, which approximates the prescribed potential function $v(x)=V(x, 0)$ with the required degree of accuracy over the range $a<x<b$.

If we take

$$V(x, y)=\text{Re}P(z),$$

---

† The approximation of a function by a finite number of terms of its Taylor series is a particular case of Weierstrass' theorem, since it applies only when the function is analytic, and the portion of the real axis over which the approximation is desired lies within the circle of convergence.

we obtain a harmonic function which is symmetrical with respect to OX, since $P(x)$ is real. It has no singularity, and varies, within the accuracy desired, as the function $v(x)$ over the interval $a<x<b$.

Since the polynomials of approximation $P(z)$ are obviously infinite in number, the number of possible potential distributions is also finite.

This proves propositions 1 and 2.

Let us note that although these possible potential distributions are very similar along ab they become very different when $|x^2+y^2|$ increases. Consider for instance two polynomials $P(x)$ and $P^*(x)$, such that for $a<x<b$

$$|f(x)-P^*(x)|<\epsilon/2\,, \quad |f(x)-P(x)|<\epsilon/2\,,$$

where ϵ is a given number, as small as desired.

On the other hand, since $\mathrm{Re}P^*(z)-\mathrm{Re}P(z)$ is a polynomial in x and y, $|\mathrm{Re}P^*(z)-\mathrm{Re}P(z)|$ becomes infinite for $x\to\infty$ or $y\to\infty$.

Moreover, the potential distributions can be made very different even in the neighbourhood of the axis of symmetry, as follows from proposition 3, whose proof is given below.

We want to find a polynomial $P(x)$ such that

$$|v(x)-P(x)|<\epsilon \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

for $a<x<b$, and

$$\mathrm{Re}P(z_j)=\mathrm{Re}P(\bar{z}_j)=V_j \quad (j=1, 2, \ldots, n) \quad . \quad . \quad . \quad . \quad (6)$$

at the points $z_j=x_j+iy_j$, $\bar{z}_j=x_{\cdot}-iy_j$ outside ab.

The function $V(x, y)=\mathrm{Re}P(z)$ will then satisfy the conditions required.

Consider the polynomial

$$P(z)=\sum_{j=1}^{n} V_j A(z)\left[\frac{1}{(x-z_j)A'(z_j)}+\frac{1}{(x-\bar{z}_j)A'(\bar{z}_j)}\right]+p(z)A(z),$$

where

$$A(z)=(z-z_1)(z-\bar{z}_1)(z-z_2)(z-\bar{z}_2)\ldots(z-z_n)(z-\bar{z}_n),$$

$$A'(z_j)=\left(\frac{dA}{dz}\right)_{z=z_j}=\left(\frac{A(z)}{z-z_j}\right)_{z=z_j},$$

and $p(z)$ is a polynomial to be determined later on.

The polynomial $P(z)$ obviously satisfies condition (6), since

$$A(z_j)=A(\bar{z}_j)=0,$$

and

$$\left[\frac{A(z)}{(z-z_j)A'(z_j)}\right]_{z=z_j}=\left[\frac{A(z)}{(z-\bar{z}_j)A'(\bar{z}_j)}\right]_{z=\bar{z}_j}=1.$$

Consider now condition (5). Let us put

$$f(x)=\frac{v(x)}{A(x)}-\sum_{j=1}^{n} V_j\left[\frac{1}{(x-z_j)A'(z_j)}+\frac{1}{(x-\bar{z}_j)A'(\bar{z}_j)}\right].$$

The function $f(x)$ is continuous along ab and is real for x real, since $A(x)$ is real, and $\frac{1}{(x-z_j)A'(z_j)}$ and $\frac{1}{(x-\bar{z}_j)A'(\bar{z}_j)}$ are conjugate.

Therefore it follows from Weierstrass' theorem that we can choose the real polynomial $p(x)$ such as

$$\left|\frac{v(x)}{A(x)}-\sum_{j=1}^{n}V_j\left[\frac{1}{(x-z_j)A'(z_j)}+\frac{1}{(x-\bar{z}_j)A'(\bar{z}_j)}\right]-p(x)\right|<\frac{\epsilon}{\text{Max}\,|A(x)|},$$

where Max $|A(x)|$ is the maximum value of $|A(x)|$ along the interval ab.

Multiplying both sides of the inequality above by $|A(x)|$ we thus have

$$\left|v(x)-\sum_{j=1}^{n}V_jA(x)\left[\frac{1}{(x-z_j)A'(z_j)}+\frac{1}{(x-\bar{z}_j)A'(\bar{z}_j)}\right]-p(x)A(x)\right|$$

$$<\epsilon\frac{|A(x)|}{\text{Max}\,|A(x)|}\leqslant\epsilon,$$

and the polynomial $P(z)$ satisfies condition (5).

It has to be noted that proposition 3 gives only a particular instance of various supplementary conditions that can be imposed on the potential function which approximates a given function along a finite portion of the axis of symmetry. Quite a wide range of different supplementary conditions could be imposed as well.

In practice the polynomials approximating a given function are usually taken to be either polynomials obtained through interpolation (Whittaker and Robinson 1940) or sums of orthogonal polynomials (Michell and Belz 1937). Trigonometric polynomials, instead of algebraic polynomials, can also be used.

Figs. 3*a* and 3*b* illustrate Proposition 2, concerning the elimination of singularities.

The potential function $v(x)$ to be obtained in the interval $-2<x<2$, $y=0$, is given as follows

$$v(x)=\frac{1}{1+x^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

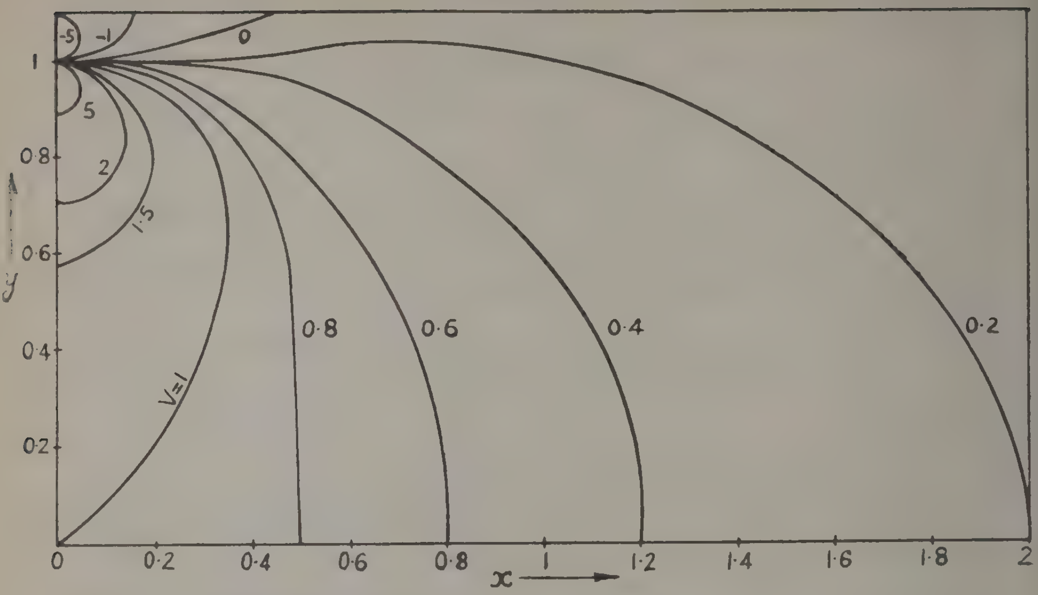
Therefore the potential function which gives exactly (7) in the interval $(-2, +2)$ is

$$V(x_1\,y)=\text{Re}\frac{1}{1+z^2}=\frac{1+x^2-y^2}{(1+x^2-y^2)^2+4x^2y^2}. \quad . \quad . \quad . \quad . \quad (8)$$

Fig. 3*a* gives the equipotential lines corresponding to (8). The point $x=0$, $y=1$ is a singular point, and in its neighbourhood $V(x, y)$ takes all the values from $-\infty$ to $+\infty$.

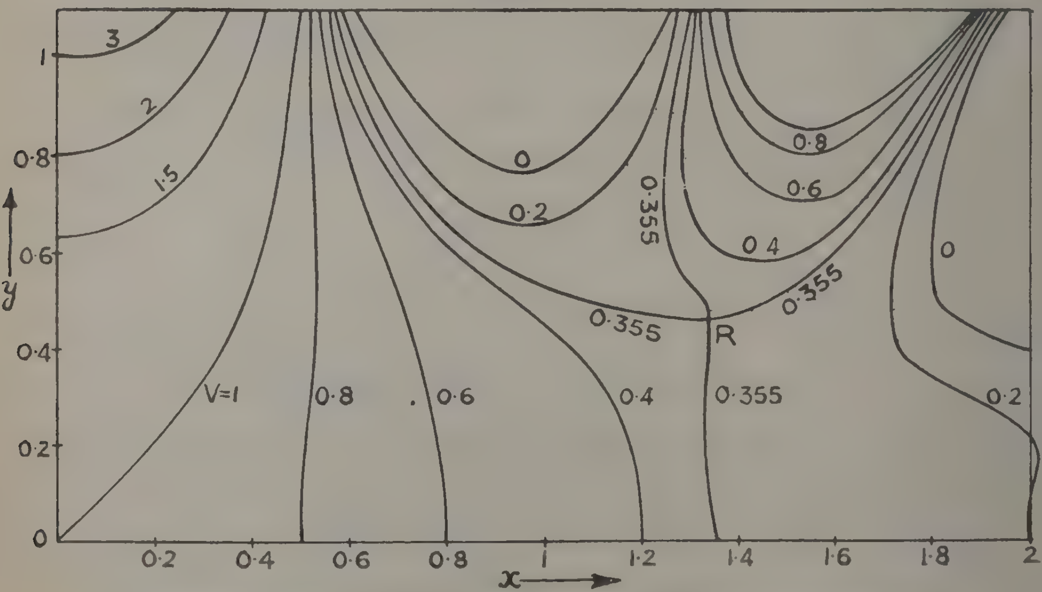
Fig. 3*b* shows an approximate solution to the problem. We wanted to obtain a polynomial $P(x)$ such that $|v(x)-P(x)|$ and $\left|\frac{dv}{dx}-\frac{dP}{dx}\right|$ are small. Therefore a polynomial of interpolation $\frac{dP}{dx}$ of 9th degree

Fig. 3*a*.



Potential distribution which corresponds to $V(x)=\frac{1}{1+x^2}$.

Fig. 3*b*.



Approximate solution. The potential along the OX axis approximates the function $\frac{1}{1+x^2}$ with an error smaller than 0·015.

has been used†, which takes the same value as $\frac{dv}{dx}$ at the points $x_j = \pm 2 \cos \frac{(2j-1)\pi}{20}$, $j=1, 2, , \ldots 5$. (As shown in Appendix I this choice of points leads to a good approximation by interpolation polynomials.)

We then obtain

$$\frac{dP}{dx} = -1{\cdot}89485x + 2{\cdot}547087x^3 - 1{\cdot}477347x^5 + 0{\cdot}383708x^7 - 0{\cdot}03636425x^9,$$

and by integration

$$P(x) = 1 - 0{\cdot}94742x^2 + 0{\cdot}63677x^4 - 0{\cdot}246224x^6 + 0{\cdot}0479635x^8 - 0{\cdot}00363642x^{10}.$$

The error limits in the interval $-2 < x < 2$, $y=0$ were found to be

$$\left|\frac{dv}{dx} - \frac{dP}{dx}\right| < 0{\cdot}05, \qquad |v(x) - P(x)| < 0{\cdot}015,$$

where

$$\text{Max}\left|\frac{dv}{dx}\right| = 0{\cdot}65, \qquad \text{Max}\,|v(x)| = 1.$$

The approximate potential distribution $V^*(x, y)$ is given by

$$\begin{aligned} V^*(x, y) &= \text{Re}P(z) \\ &= P(x) - \frac{y^2}{2!}\frac{d^2P}{dx^2} + \frac{y^4}{4!}\frac{d^4P}{dx^4} + \ldots - \frac{y^{10}}{10!}\frac{d^{10}P}{dx^{10}}. \end{aligned}$$

The last line is an application of formula (3).

Fig. 3*b* shows the equipotential lines corresponding to $V^*(x, y)$. They are very similar to the lines of fig. 3*a* in the strip $0 < y < 0{\cdot}10$. They become very different for values of y larger than 0·30. Of course $V^*(x, y)$ has no singularities, and the branch point R is a point where the electric field is zero.

§ 4. CASE OF ROTATIONAL SYMMETRY.

The same general results are obtained in this case, if formulæ (1) (2) (3) are replaced by

$$\frac{\partial^2 V}{\partial z^2} + \frac{1}{r} \cdot \frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial r^2} = 0,$$

$$V(r, z) = \frac{1}{\pi}\int_0^\pi v(z + ir\cos\phi)\, d\phi,$$

$$V(r, z) = v(z) - \frac{r^2}{2^2} \cdot \frac{d^2v}{dz^2} + \frac{r^4}{2^2 4^2} \cdot \frac{d^4v}{dz^4} + \ldots + \frac{(-1)^n r^{2n}}{2^2 4^2 \ldots (2n)^2} \cdot \frac{d^{2n}v}{dz^{2n}} + \ldots, \qquad (9)$$

where r and z are cylindrical coordinates.

---

† If $\left|\frac{dv}{dx} - \frac{dP}{dx}\right| < \epsilon$ we have $\left|v(x) - P(x)\right| = \left|\int_0^x \left(\frac{dv}{dx} - \frac{dP}{dx}\right) dz\right| = \epsilon x$. Hence a good approximation of $\frac{dv}{dx}$ leads to a good approximation of $v(x)$ (if x is not too large), whereas the converse is not true.

Two interesting examples concerning this case are given by N. Plass (1942), who considers the equipotential lines which produce the axial potentials $v(z)=\mathrm{A}\exp(-z^2/2)$ (Scherzer's lens of minimum aberration), and $v(z)=\mathrm{A}\tanh z$. Plass uses formula (9) where the functions of z are also expanded in Taylor series, and retains only a finite number of terms. N. Plass considers his method to be valid for small values of r only, because he implicitly assumes that the equipotential lines must have approximately the shape of those derived from the rigorous solution. It is seen from the present paper that such an assumption is not necessary, and therefore the results obtained by N. Plass can be extended to large values of r. It has also to be noted that a better approximation is usually obtained when $v(x)$ is expressed by an interpolation polynomial (see Appendix), or orthogonal polynomials, instead of using a finite Taylor series.

ACKNOWLEDGMENTS.

The author is indebted to Dr. D. Gabor of the Imperial College for suggesting this research, and wishes to express appreciation for his helpful interest in this work.

Thanks are also due to the Directors of Mullard Radio Valve Co. for permission to publish this paper.

APPENDIX.

ABOUT THE CHOICE OF INTERPOLATION POINTS.

Known and useful results based on an extremal property of Tchebycheff's polynomials are expounded here, since the works dealing with them are not easily available.

It is well known that if $\mathrm{P}_{n-1}(x)$ is the polynomial of degree $n-1$ which takes the same values as $f(x)$ at points x_j ($j=1, 2, \ldots, n$) in the interval ab, then in this interval (Whittaker and Robinson 1940),

$$|f(x)-\mathrm{P}_{n-1}(x)|\leqslant\frac{1}{n!}\left|\frac{d^nf}{dx^n}\right|\underset{a<x<b}{\mathrm{Max}}|\mathrm{A}_n(x)|, \quad . \quad . \quad . \quad . \quad (10)$$

where

$$\mathrm{A}_n(x)=(x-x_1)(x-x_2)\ldots(x-x_n), \quad . \quad . \quad . \quad . \quad (11)$$

and where it is assumed that $\frac{d^nf}{dx^n}$ is continuous.

To obtain the smallest upper limit for $|f(x)-\mathrm{P}_{n-1}(x)|$ as given by (10), the points of interpolation $x_1, x_2, \ldots, x_n$ must be such as to ensure the smallest value of $\underset{a<x<b}{\mathrm{Max}}|\mathrm{A}_n(x)|$.

The interval of interpolation ab may be assumed to correspond to $-1 \leqslant x \leqslant 1$, which can always be the case after a suitable change of variables. Then the smallest value of $\underset{-1<x<1}{\text{Max}} |A_n(x)|$ is obtained if the points $x_1, x_2, \ldots, x_n$ are the zeros of the Tchebycheff polynomial of degree n, $T_n(x)$, or more precisely if

$$A_n(x) \equiv T_n(x).$$

This theorem is due to Tchebycheff (1899), and a short proof of it will be given here.

The Tchebycheff polynomial $T_n(x)$ is given by

$$\begin{aligned} T_n(x) &= \frac{\cos(n \cos^{-1} x)}{2^{n-1}} \\ &= x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_0{}^{*}. \qquad (12) \end{aligned}$$

Its zeros are in the interval $(-1, +1)$, and are given by $x_j = \cos \frac{(2j-1)\pi}{2n}$, $j = 1, 2, \ldots, n$. Therefore $T_n(x)$ can be put in the form (11).

On the other hand, for $A_n(x)$ different from $T_n(x)$, we have

$$\underset{-1 \leqslant x \leqslant 1}{\text{Max}} |A_n(x)| > \underset{-1 \leqslant x \leqslant 1}{\text{Max}} |T_n(x)|. \;.\;.\;.\;.\;.\;.\;. \qquad (13)$$

This can be derived as follows.

From (12) it is seen that

$$\text{Max}\, |T_n(x)| = \frac{1}{2^{n-1}}$$

and

$$T_n\left(\cos \frac{m\pi}{n}\right) = \frac{(-1)^m}{2^{n-1}}, \quad m = 0, 1, 2, , , , , n.$$

Let us now suppose that

$$\underset{-1 \leqslant x \leqslant 1}{\text{Max}} |A_n(x)| < \underset{-1 \leqslant x \leqslant 1}{\text{Max}} |T_n(x)|.$$

Then the polynomial $T_n(x) - A_n(x)$ will be positive at the points $x_m = \cos \frac{m\pi}{n}$ corresponding to m even, and negative at the points x_m

---

[*] The coefficients a_j are determined by putting

$$x = \cos\theta$$

$$\begin{aligned} T_n(x) = \frac{\cos n\theta}{2^{n-1}} &= \frac{1}{2^n}[\exp(in\theta) + \exp(-in\theta)] \\ &= \frac{1}{2^n}[(\cos\theta + i\sin\theta)^n + (\cos\theta - i\sin\theta)^n] \\ &= \frac{1}{2^n}\,[(x + \sqrt{(x^2-1})^n + (x - \sqrt{(x^2-1})^n]. \end{aligned}$$

corresponding to m odd, and hence $T_n(x)-A_n(x)$ will change its sign n times. It must therefore have n zeros, which is impossible since $T_n(x)-A_n(x)$ is of $n-1$th degree.

For a similar reason it is not possible to have

$$\underset{-1\leqslant x\leqslant 1}{\text{Max}}|A_n(x)| = \underset{-1\leqslant x\leqslant 1}{\text{Max}}|T_n(x)|.$$

Hence (13) has been proved, and the upper limit of $|f(x)-P_{n-1}(x)|$ as given by (10) is minimum when $A_n(x)\equiv T_n(x)$.

In this case

$$|f(x)-P_{n-1}(x)|\leqslant\frac{1}{n!}\frac{1}{2^{n-1}}\underset{-1\leqslant x\leqslant 1}{\text{Max}}\left|\frac{d^nf}{dx^n}\right|,$$

and it is seen that the upper limit is much lower than when $P_{n-1}(x)$ is a finite Taylor series $\tau_{n-1}(x)$, in which case we have

$$|f(x)-\tau_{n-1}(x)|\leqslant\frac{1}{n!}\underset{-1\leqslant x\leqslant 1}{\text{Max}}\left|\frac{d^nf}{dx^n}\right|.$$

REFERENCES.

COSSLETT, V. E., 1946, *Introduction to Electron Optics*, pp. 25, 128. Oxford:

GOURSAT, E., 1927, *Cours d'Analyse Mathématique*, **1,** 498–500. Paris.

MICHELL, J. H., and BELZ, M. M., 1937, *The Elements of Mathematical Analysis*, Vol. II., Chapter XVII.

PLASS, J. N., 1942, "Electrostatic Electron Lenses with a Minimum of Spherical Abberration", *Journal of Applied Physics*, **13,** 49–55.

TCHEBYCHEFF, P. L., 1899, *Oeuvres*, Vol. I. St. Petersburg.

WHITTAKER, E., and ROBINSON, G., 1940, *The Calculus of Observations*. Chapters I and II.

XX. *Properties of Slow Electrons in Polar Materials**.

By H. Fröhlich, H. Pelzer† and S. Zienau,
Department of Theoretical Physics, The University, Liverpool‡.

[Received November 25, 1949.]

Summary.

Using a variational method we have investigated the properties of the lowest energy levels in a range $\hbar\omega$ above the ground level of the system consisting of an electron and a continuous dielectric medium. The latter is supposed to have a single vibrational frequency $\omega/2\pi$ for long longitudinal polarization waves. The interaction between the electron and the medium then depends on three parameters, the static dielectric constant ϵ, the optical refractive index $\epsilon_{\infty}^{\frac{1}{2}}$, and ω. The replacement of a crystalline lattice by a continuum is a good approximation if the length $b=(h/2m\omega)^{\frac{1}{2}}$ is large compared with the lattice distance, a condition which is usually fulfilled.

We find that the energy of the ground level may be considerably (compared with $\hbar\omega$) below the energy which the system would have in the absence of interaction. The electron can be found with equal probability at any point in the medium. The average polarization of the medium therefore vanishes at any point. This does not hold, however, for the average polarization at a given distance from the electron. This quantity varies at large distances as the polarization of a point charge, but shows deviations below distances of the order b. The energy of interaction depends very little on the average velocity of the electron. Slow electrons, therefore, behave very similarly to free electrons (§4). It follows then that self trapping in the lattice—a suggestion which has often been discussed—does not exist (§5).

Modified forms of previous formulæ of the mean free path of electrons are given in §6. It is shown, however, that the validity of the whole method used at present to calculate mean free paths requires further investigation.

§1. Introduction.

Through combined theoretical and experimental work during the past twenty years it has been established that slow electrons in periodic crystal lattices behave in many respects like free electrons; the main influence of the periodic field is to replace the electronic mass by a slightly different effective mass. The effect of deviations from the periodicity, *e. g.* by thermal motion or through lattice defects can then be described as leading to scattering of otherwise free electrons. It might seem that ionic crystals differ in this respect from others. For in these crystals the motion of ions has a resonance at frequencies which makes their displacement by moving charges strongly dependent on the velocity of the latter. In fact it has been suggested that in view of their polarization an electron might be trapped near any point in the lattice ; and that a certain energy would have to be supplied to transfer it into a state in which it can move more freely. Our calculations show that this is not the case.

---

* Based on Report L/T 221 of the British Electrical and Allied Industries Research Association (E.R.A.).

† On the research staff of the British Electrical and Allied Industries Research Association.

‡ Communicated by the Authors.

Electronic motion in ionic crystals is of importance for a considerable number of physical processes. In the present paper, therefore, we have made a systematic study of the properties of the lowest energy levels of an electron, assuming it to be free except for reaction forces due to the enforced displacements of ions.

To obtain a qualitative idea of the interaction between an electron and the polarization produced by it let us first consider the behaviour of a moving point charge e [*cf.* Fröhlich and Pelzer (1950)] and then discuss the modifications required if the point charge is replaced by an electron. The polarization $\mathbf{P}$ in an ionic crystal can be considered as super-position of two components [*cf. e. g.* Fröhlich (1949)] : (i) the optical polarization $\mathbf{P}_0$ which is due to the displacement of bound electrons of the lattice with a resonance frequency in the optical or ultra-violet region ; (ii) the infra-red polarization $\mathbf{P}_i$, whose resonance frequency $\omega/2\pi$ lies in the infra-red and is connected with the displacement of ions. For the slow velocities v of the point charge in which we are interested the optical polarization is always excited to its static value whatever the velocity v. The interaction of $\mathbf{P}_0$ with the point charge (or the electron) therefore is of no interest to us as it does not influence the energy difference between states of different velocity. The infra-red polarization behaves differently in this respect because of the lower resonance frequency of ionic oscillations. At large distances from the point charge (position $\mathbf{r}_0$) the polarization $\mathbf{P}_{ir}$ is of course the same as in the static case. It is thus obtained by subtracting the optical polarization $\mathbf{P}_0$ from the total polarization $\mathbf{P}$, and can be derived from an electric potential $\Phi(\mathbf{r})$ by the equation

$$4\pi \boldsymbol{P}_{ir}(\mathbf{r}) = \operatorname{grad} \Phi(\mathbf{r}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.1)$$

where according to electrostatics

$$\Phi(\mathbf{r}) = -\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon}\right) \frac{e}{|\mathbf{r}-\mathbf{r}_0|}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.2)$$

Here $\boldsymbol{r}$ is any point at sufficient distance from $\mathbf{r}_0$, ϵ is the static dielectric constant and ϵ_∞ is the dielectric constant at high frequencies where the ionic motion cannot follow. $\epsilon_\infty^{\frac{1}{2}}$ is the optical refractive index. An ion is displaced as in the static case only if its distance from the approaching point charge is sufficiently larger than v/ω. The reason is that it takes a time of the order $1/\omega$ to displace an ion. Below this distance the point charge passes so quickly that its action on the ion can be considered as a shock, exciting oscillation after it has passed. In this case the potential at the position of the point charge is obtained by replacing $|\mathbf{r}-\mathbf{r}_0|$ in (1.2) by v/ω; multiplication by e leads to the energy of interaction. To obtain the net interaction energy we must add the energy required to polarize the lattice. For harmonically bound ions this reduces the interaction energy to about half of the above value. Its order of magnitude is thus

$$-\frac{1}{2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\frac{e^2\omega}{v}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.3)$$

If the point charge is replaced by an electron the uncertainty relation requires that we cannot confine it to a small region in space without a considerable uncertainty in its velocity. Let this region be of linear dimension a around a point $\mathbf{r}_0$. Thus at a large distance from $\mathbf{r}_0$ the potential is still given by expression (1.2). At short distances there are now two competitive effects which prevent the potential from falling below a certain value. Considering first the uncertainty in position only, the magnitude of the lowest value of the potential would be obtained by replacing $|\mathbf{r}-\mathbf{r}_0|$ in (1.2) by a. On the other hand, considering the effect of the velocity only, $|\mathbf{r}-\mathbf{r}_0|$ must be replaced by v/ω. We should expect that it is the mechanism which leads to the larger critical distance which decides the order of magnitude of the potential inside the region to which the electron is confined. The two distances are not independent, however, because according to the uncertainty relation $\Delta v \simeq h/ma$ and Δv decides the order of magnitude of v if the point $\mathbf{r}_0$ is either at rest, or moving only slowly. A minimum for the critical distance exists, therefore, being the lower value of the two quantities v/ω and h/mv. It is hence of the order

$$b=\left(\frac{\hbar}{2m\omega}\right)^{\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.4)$$

a length which will be of importance later on*. It now follows that a minimum value for the negative interaction energy should exist whose magnitude is of the order

$$W=\frac{1}{2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\frac{e^2}{b}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.5)$$

It is suggestive to imagine the electron to be trapped by the potential field due to the polarization induced by itself, as was first done by Landau (1933) and later on followed up by others. The electron cloud is supposed to be centred around a fixed point in the medium: the exact distribution depends on the interaction with the polarization of the medium which in turn depends on the distribution of charge density. By making polarization and density distribution self-consistent a solution is obtained which is supposed to have the following properties: (i.) it constitutes the lowest energy level of the system; (ii.) its mobility is so low and its effective mass so high that the electron can be considered as trapped; (iii.) by supplying a minimum energy the electron can be liberated into an unbound state, similar to the case that an electron is bound (trapped) by a permanent potential. These properties, if true, would be of great importance for the behaviour of electrons in ionic crystals. No attempts have been made so far to prove any of the three contentions. They are all objectionable, and we shall find in fact that none are valid (§5).

---

* The exact numerical value has been chosen for later convenience.

§ 2. CALCULATIONS; GENERAL METHOD.

Following the treatment in a previous paper [Fröhlich, (1937) quoted as (I.)] we shall investigate properties of a system consisting of an electron and a continuous dielectric medium possessing a single infra-red vibrational frequency. The electron is treated as free apart from its interaction with this medium. Any effect of a periodic lattice field might be taken into account by the use of an effective electronic mass. In the previous paper the scattering of a relatively fast (several eV.) electron was considered by treating its interaction with the medium as a small perturbation. At present we shall be mainly interested in a calculation of the position and properties of the ground level of the system, and of the subsequent levels in a range of the order of 1/10 eV. For this purpose we require the Hamiltonian H consisting essentially of three parts: (i) the energy of the dielectric, described by its polarization; (ii) the electron's energy of interaction with the polarization; (iii) the kinetic energy of the electron. As indicated in the introduction the energy of interaction between the electron and the optical polarization is the same for all states in which we are interested at present. It thus leads to a constant term in the Hamiltonian which we shall omit. The state of the dielectric is then entirely described by the space and time dependence of the infra-red polarization $\mathbf{P}_{ir}$. It is convenient to develop $\mathbf{P}_{ir}$ into a Fourier series of polarization waves using a term introduced on a previous occasion (I.). Thus we can write

$$\mathbf{P}_{ir}=\frac{1}{V^{\frac{1}{2}}}\sum_{\mathbf{w}}(\mathbf{a}_{\mathbf{w}}\exp\{i\mathbf{wr}\}+\text{c.c.}), \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

where V denotes the volume in which we are interested and c.c. stands for the complex conjugate of the preceding expression. In the absence of external forces (*e. g.* due to the electron) the $\mathbf{a}_{\mathbf{w}}$ depend periodically on time. As in (I.) longitudinal waves only are of interest because transverse waves do not interact with an electron. In contrast to (I.) we shall be mainly interested in polarization waves which are long compared with the lattice distance. It is for this reason that the dielectric can be treated as a continuum. The frequency $\omega/2\pi$ of all waves can then be considered as independent of wavelength. This frequency differs however from that of very short waves; its value will be discussed in §4.

It is useful to introduce the electric potential $\Phi(\mathbf{r})$ from which the polarization $\mathbf{P}_{ir}$ can be derived according to (1.1) because then the energy of interaction is simply given by

$$H_{\text{int}}=e\Phi(\mathbf{r}_e), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

where $\mathbf{r}_e$ is the position of the electron. Introducing two new sets of coordinates $X_\mathbf{w}$ $Y_\mathbf{w}$ and a constant M by

$$X_\mathbf{w}=a_\mathbf{w}+a^*_\mathbf{w}, \quad Y_\mathbf{w}/M\omega=-i(a_\mathbf{w}-a^*_\mathbf{w}), \quad . \quad . \quad . \quad . \quad (2.3)$$

we obtain from (1.1) and (2.1)

$$\Phi(\mathbf{r})=\frac{4\pi}{V^{\frac{1}{2}}}\sum_\mathbf{w}\frac{1}{w}\left(X_\mathbf{w}\sin\mathbf{wr}+\frac{Y_\mathbf{w}}{M\omega}\cos\mathbf{wr}\right). \quad . \quad . \quad . \quad (2.4)$$

Clearly without a field $Y_\mathbf{w}=M\dot{X}_\mathbf{w}$. The constant M was introduced so that formally the medium is represented by a set of harmonic oscillators with mass M, displacement $X_\mathbf{w}$ and momentum $Y_\mathbf{w}$. We shall choose the constant M in such a way that the oscillator energy is given by

$$H_{\text{osc}}=\sum_\mathbf{w}\frac{1}{2}\left(M\omega^2X^2_\mathbf{w}+\frac{Y^2_\mathbf{w}}{M}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.5)$$

It should be clear, however, that neither has M the dimension of a mass nor X and Y those of length and momentum respectively.

Denoting by $\mathbf{p}$ the momentum of the electron the total Hamiltonian is given by

$$H=H_{\text{osc}}+e\Phi(\mathbf{r}_e)+\frac{p^2}{2m}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.6)$$

The $X_\mathbf{w}$ and $Y_\mathbf{w}$ and $\mathbf{r}_e$ and $\mathbf{p}$ can then be treated as sets of canonical conjugate coordinates.

We might determine M from the known energy of a simple (*e. g.* homogeneous) displacement. We shall instead calculate the potential due to a point charge at $\mathbf{r}_0$ and determine M by comparison with the known result (1.2). This will give us an opportunity of demonstrating the method of handling (2.6) in a simple case. The replacement of the electron by a point charge implies that the kinetic energy term $p^2/2m$ must be removed from (2.6) and $\mathbf{r}_e$ must be replaced by the constant $\mathbf{r}_0$. Then if we introduce the constants

$$\overline{X}_\mathbf{w}=-\frac{4\pi e}{V^{\frac{1}{2}}}\frac{1}{w}\frac{\sin\mathbf{wr}_0}{M\omega^2}, \qquad \overline{Y}_\mathbf{w}=-\frac{4\pi e}{V^{\frac{1}{2}}}\frac{1}{w}\frac{\cos\mathbf{wr}_0}{\omega}, \quad . \quad . \quad (2.7)$$

clearly

$$H=\sum_\mathbf{w}\left(H_\mathbf{w}-\tfrac{1}{2}M\omega^2\overline{X}^2_\mathbf{w}-\frac{1}{2}\frac{\overline{Y}^2_\mathbf{w}}{M}\right), \quad . \quad . \quad . \quad . \quad (2.8)$$

where

$$H_\mathbf{w}=\tfrac{1}{2}M\omega^2(X_\mathbf{w}-\overline{X}_\mathbf{w})^2+\frac{1}{2M}(Y_\mathbf{w}-\overline{Y}_\mathbf{w})^2. \quad . \quad . \quad . \quad (2.9)$$

By this transformation our Hamiltonian has been separated into a constant term and a set of Hamiltonians of harmonic oscillators with coordinates $X_{\mathbf{w}}-\overline{X}_{\mathbf{w}}$ and momenta $Y_{\mathbf{w}}-\overline{Y}_{\mathbf{w}}$.

Quantum-mechanically, $\overline{X}$ and $\overline{Y}$ are the expectation values of X and Y; classically they are the averages over one period. The expectation value for $\Phi(\mathbf{r})$, using equation (2.4) is thus given by

$$\Phi(\mathbf{r}-\mathbf{r}_0)=-\frac{16\pi^2 e}{VM\omega^2}\sum_{\mathbf{w}}\frac{\cos(\mathbf{w}.\mathbf{r}-\mathbf{r}_0)}{\mathbf{w}^2}. \quad . \quad . \quad . \quad (2.10)$$

On this, and on all later occasions, the sum will be transformed into an integral in the usual way by counting the number of values which $\mathbf{w}$ can have in a small range $w^2\,dw\,d\Omega$ on the assumption of a periodic boundary condition. Thus

$$\sum_{\mathbf{w}}(\quad)\to\frac{V}{(2\pi)^3}\int w^2\,dw\,d\Omega(\quad); \quad . \quad . \quad . \quad . \quad . \quad (2.11)$$

where $d\Omega$ is the element of the solid angle in $\mathbf{w}$-space. The integration goes over all $\mathbf{w}$-space*. The result agrees with (1.2) if we determine M from

$$\frac{1}{M\omega^2}=\frac{1}{4\pi}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.12)$$

Unfortunately when replacing the point charge by an electron this simple method can no longer be used and a suitable approximative method has to be found. The electron as well as the oscillators will be treated quantum-mechanically.

Let us first express the Hamiltonian in a useful matrix representation. We shall make use of the complete orthogonal set of wave functions $\Psi_0(\mathbf{K};\ n_1, n_1, \ldots n_{\mathbf{w}}, \ldots)$ which diagonalizes H when the interaction term $e\Phi(\mathbf{r}_e)$ in (2.6) is neglected. These functions have previously been discussed in (I.). They can be written as

$$\Psi_0(\mathbf{K};\ n_1,\ldots,n_{\mathbf{w}},\ldots)=\frac{\exp\{i\mathbf{K}\mathbf{r}\}}{V^{\frac{1}{2}}}\prod_{\mathbf{w}}\chi_{\mathbf{w}}(X_{\mathbf{w}}, n_{\mathbf{w}}), \quad . \quad (2.13)$$

where $\chi_{\mathbf{w}}(n_{\mathbf{w}})$ is the wave function of a harmonic oscillator with coordinate $X_{\mathbf{w}}$ in its $n_{\mathbf{w}}$th excited state. In this representation, the diagonal elements of the interaction, H_{int}, vanish. Thus the diagonal elements of H are given by

$$E_0(\mathbf{K};\ n_1,\ldots)=\frac{\hbar^2\mathbf{K}^2}{2m}+\hbar\omega\sum_{\mathbf{w}}(n_{\mathbf{w}}+\tfrac{1}{2}), \quad . \quad . \quad . \quad (2.14)$$

corresponding to the energy when interaction is neglected. The non-diagonal elements vanish except for those corresponding to the absorption or emission of one quantum $\hbar\omega$ by the electron. If $\mathbf{w}$ is the wave number of this quantum then they will be denoted by $(e\Phi)_{\mathbf{KK'w}}$ to show that only the interaction energy gives a contribution. Their value is found in a straight forward way similar as in (I., p. 236) by making use of (2.4) and of the well known matrix elements of the coordinate and the momentum of a harmonic oscillator. Introducing the value (2.12) for M we find

---

* In (I.) it was necessary to introduce an upper limit w_0 of w. This is unnecessary at present as the main contributions to all integrals required will be found to arise from much smaller values of w.

$$(e\Phi)_{\mathbf{KK'w}}=(e\Phi)_{\mathbf{w}}\left\{\begin{array}{l}(1+n_{\mathbf{w}})^{\frac{1}{2}}\\ -n_{\mathbf{w}}^{\frac{1}{2}}\end{array}\right. , \qquad (2.15)$$

where

$$(e\Phi)_{\mathbf{w}}=\frac{4\pi ie}{V^{\frac{1}{2}}}\frac{1}{w}\left[\frac{\hbar\omega}{8\pi}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\right]^{\frac{1}{2}}. \qquad (2.16)$$

Here $(1+n_{\mathbf{w}})^{\frac{1}{2}}$ holds in the case of an emission of a quantum $\hbar\omega$ with wave number $\mathbf{w}$ by the electron, provided

$$\mathbf{K}'=\mathbf{K}-\mathbf{w}. \qquad (2.17)$$

For absorption $n_{\mathbf{w}}^{\frac{1}{2}}$ must be used instead, and

$$\mathbf{K}'=\mathbf{K}+\mathbf{w}. \qquad (2.18)$$

All non-diagonal elements for which neither of these conditions is fulfilled vanish.

We shall derive our solutions from those of ordinary second order perturbation theory, which starts in zero order with an electron with energy smaller than $\hbar\omega$ and with all $n_{\mathbf{w}}=0$. Such an electron cannot emit a real quantum. It can, however, emit, and reabsorb virtually single quanta provided (2.17) and (2.18) are fulfilled respectively. The opposite sequence, first absorption and then emission is impossible as initially all $n_{\mathbf{w}}$ vanish. After previous emission, just one $n_{\mathbf{w}}=1$ is present however. Apart from the sign, both matrix elements are then equal, putting in (2.15) $n_{\mathbf{w}}=0$ and $n_{\mathbf{w}}=1$ respectively. They are then given by $\pm(e\Phi)_{\mathbf{w}}$, equation (2.16). The energies in zero order in the initial state $E_0(K)$ and in the intermediate state $E_0(\mathbf{K},\mathbf{w})$ respectively are given by [making use of (2.14) and (2.17)]

$$E_0(K)=\frac{\hbar^2K^2}{2m}, \qquad E_0(\mathbf{K},\mathbf{w})=\frac{\hbar^2(\mathbf{K}-\mathbf{w})^2}{2m}+\hbar\omega, \qquad (2.19)$$

if we omit the zero point energy of the oscillators, $\Sigma\hbar\omega/2$. This is admissible as we are only concerned with energy differences. Thus, according to the well known formula, the perturbed energy is given by

$$\begin{aligned}E(K)&=E_0(K)-\sum_{\mathbf{w}}\frac{|(e\Phi)_{\mathbf{w}}|^2}{E_0(\mathbf{K},\mathbf{w})-E_0(\mathrm{K})}\\ &=\frac{\hbar^2K^2}{2m}-\frac{2m}{\hbar^2}\sum_{\mathbf{w}}\frac{|(e\Phi)_{\mathbf{w}}|^2}{(\mathbf{K}-\mathbf{w})^2+u^2-K^2}, \qquad (2.20)\end{aligned}$$

where we have introduced a constant u by (*cf.* 1.4)

$$u=\frac{1}{b}, \quad i.e.\ \frac{\hbar^2u^2}{2m}=\hbar\omega. \qquad (2.21)$$

The corresponding wave function becomes

$$\Psi(\mathbf{K})=\Psi_0(\mathbf{K}\,;0,0,\ldots)+\sum_{\mathbf{w}} c_{\mathbf{w}}\Psi_0(\mathbf{K}-\mathbf{w}\,;1_{\mathbf{w}}), \quad . \quad . \quad (2.22)$$

where

$$c_{\mathbf{w}}=-\frac{(e\Phi)_{\mathbf{w}}}{E_0(\mathbf{K},\,\mathbf{w})-E_0(K)}=-\frac{2m}{\hbar^2}\,\frac{(e\Phi)_{\mathbf{w}}}{(\mathbf{K}-\mathbf{w})^2+u^2-K^2}. \quad . \quad (2.23)$$

It consists of the sum of wave functions of our set having the same momentum $\hbar\mathbf{K}$ with appropriate factors. The first term corresponds to the case in which no quantum is excited, the remaining functions describe states with one excited quantum each.

We shall generalize the above expressions by continuing to use a wave function of the type (2.22) because without specializing the $c_{\mathbf{w}}$ this is the most general wave function with total momentum $\hbar\mathbf{K}$ containing functions with none or one excited quantum. It is to be hoped that this will suffice to describe the lowest energy levels, except in the cases of very strong interaction. The coefficients $c_{\mathbf{w}}$, and the energy $E(K)$ will be determined by minimizing the expectation value of the energy, with respect to $c_{\mathbf{w}}$. Thus

$$E(K)=\int\Psi^*H\Psi\Big/\int\Psi^*\Psi. \quad . \quad . \quad . \quad . \quad . \quad (2.24)$$

The integration goes over all variables in Ψ. Making use of the matrix elements of H discussed above, we find after multiplication by $\int\Psi^*\Psi=1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2$, (the latter holds in view of the orthogonality and normalization of the functions Ψ_0)

$$\Big(1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2\Big)E(K)=E_0(K)+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2E_0(\mathbf{K},\,\mathbf{w})+\sum_{\mathbf{w}}\big(c_{\mathbf{w}}(e\Phi)_{\mathbf{w}}+\text{c.c.}\big). \quad (2.25)$$

Hence by varying with respect to $c_{\mathbf{w}}$ and $c^*_{\mathbf{w}}$ we obtain using (2.19) and (2.21)

$$c_{\mathbf{w}}=-\frac{(e\Phi)_{\mathbf{w}}}{E_0(\mathbf{K},\,\mathbf{w})-E(K)}=-\frac{(e\Phi)_{\mathbf{w}}}{\hbar^2[(\mathbf{K}-\mathbf{w})^2+u^2]/2m-E(K)} \quad (2.26)$$

and

$$\begin{aligned}E(K)&=E_0(K)-\sum_{\mathbf{w}}\frac{|(e\Phi)_{\mathbf{w}}|^2}{E_0(\mathbf{K},\,\mathbf{w})-E(K)}\\&=\frac{\hbar^2K^2}{2m}-\frac{2m}{\hbar^2}\sum_{\mathbf{w}}\frac{|(e\Phi)_{\mathbf{w}}|^2}{(\mathbf{K}-\mathbf{w})^2+u^2-2mE(K)/\hbar^2}. \quad . \quad . \quad (2.27)\end{aligned}$$

Both expressions (2.27) and (2.26) turn into the corresponding ones (2.22) and (2.20) in the perturbation theory if the energy perturbation is sufficiently small. We also notice that the normalization integral

$$J=\int\Psi^*\Psi=1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.28)$$

must differ very little from unity if perturbation theory holds. This is not required in the case of the variational method.

Finally we should like to mention that equation (2.27) for the lowest energy level with momentum $\hbar\mathbf{K}$ can also be obtained by diagonalizing the matrix representation of H with respect to states containing either none or one quantum.

§3. ENERGY, VELOCITY AND OTHER PROPERTIES.

The above expressions for the energy and the wave functions contain $\mathbf{K}$ as a parameter. The quantity $\hbar\mathbf{K}$ represents the momentum of the total system, electron+polarization waves, as can easily be verified. In fact each of the functions Ψ_0 appearing in (2.22) has the same momentum $\hbar\mathbf{K}$.

To obtain the energy in an explicit form it is useful to introduce the following dimensionless quantities [*cf.* (1.5), (1.4) and (2.21)]

$$x=\frac{K}{u}, \qquad \alpha=\frac{W}{\hbar\omega}=\frac{1}{2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\frac{e^2u}{\hbar\omega}, \qquad y=\frac{E_0(K)-E(K)}{\hbar\omega}. \quad (3.1)$$

The energy can then be written as

$$E(K)=\hbar\omega(x^2-y) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.2)$$

if use is made of equation (2.19).

We then find from (2.27) by introducing from (2.16) and using (2.11), (2.21), (1.5) and (3.1),

$$\left.\begin{aligned} y&=\frac{\alpha u}{2\pi^2}\int_0^\infty dw\int\frac{d\Omega}{(\mathbf{K}-\mathbf{w})^2+u^2(1+y)-K^2}\\ &=\frac{\alpha}{x}\sin^{-1}\frac{x}{(1+y)^{\frac{1}{2}}} \qquad \text{if} \quad 1+y>x^2\\ &=\frac{\alpha}{x}\frac{\pi}{2} \qquad\qquad\qquad\quad \text{if} \quad 0<1+y<x^2\\ &=0 \qquad\qquad\qquad\qquad\quad \text{if} \quad 1+y<0. \end{aligned}\right\} \quad . \quad . \quad . \quad (3.3)$$

We shall restrict ourselves mainly to the case $1+y>x^2$ which as we shall see corresponds to the lowest range of energies. In this case the denominator in the above integral does not vanish for any real w. In the other cases the values given above are principal values. In a more convenient form we thus have

$$\frac{x}{(1+y)^{\frac{1}{2}}}=\sin\frac{yx}{\alpha} \quad \text{if} \quad 1+y>x^2. \quad . \quad . \quad . \quad . \quad . \quad (3.4)$$

In particular, if y_0 denotes the value of y for $x=0$,

$$y_0(1+y_0)^{\frac{1}{2}}=\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.5)$$

From this expression we see that α acts as an interaction parameter. If $\alpha \ll 1$, then $y \simeq \alpha$, a result which is identical with that obtained by perturbation theory.

Near $x=0$, the quantity y increases quadratically with x. By development one obtains

$$y-y_0 \simeq \frac{y_0}{2+3y_0}\frac{x^2}{3} \quad \text{if} \quad x^2 \ll 1. \quad . \quad . \quad . \quad . \quad . \quad (3.6)$$

The upper limit of this range is reached when $\sin(yx/\alpha)$ in (3.4) becomes equal to unity, *i. e.* at values x_1 and y_1 satisfying

$$\frac{x_1}{(1+y_1)^{\frac{1}{2}}}=1, \qquad x_1 y_1 = \alpha\frac{\pi}{2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.7)$$

Hence

$$x_1^2-y_1=1, \qquad x_1(x_1^2-1)=\alpha\frac{\pi}{2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.8)$$

showing that $x_1>1$, and with (3.2), that the energy at this upper limit is equal to $\hbar\omega$. Also developing near x_1, y_1

$$(y-x^2)-(y_1-x_1^2) \simeq (3x_1^2-1)^2\frac{x_1^2}{\alpha^2}(x-x_1)^2 \quad \text{if} \quad (x-x_1)^2 \ll x_1. \quad (3.9)$$

Turning now to the average velocity of the electron we shall prove that

$$v_x=\frac{1}{\hbar}\frac{\partial E(K)}{\partial K_x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.10)$$

where v_x represents the X-component. By definition,

$$v_x=\frac{\hbar}{mi}\int \Psi^*\frac{\partial \Psi}{\partial x} \Big/ \int \Psi^*\Psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.11)$$

or introducing from (2.22) using the orthogonality of the functions Ψ_0,

$$\begin{aligned}\frac{\hbar}{m}v_x &= \Big(K_x+\sum_{\mathbf{w}}(K_x-w_x)|c_{\mathbf{w}}|^2\Big)\Big/\Big(1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2\Big) \\ &= K_x-\sum_{\mathbf{w}} w_x|c_{\mathbf{w}}|^2\Big/\Big(1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2\Big). \quad . \quad . \quad . \quad . \quad . \quad (3.12)\end{aligned}$$

Now

$$w_x=w(\cos\theta\cos\vartheta-\sin\theta\sin\vartheta\cos\phi),$$

if θ is the angle between $\mathbf{w}$ and $\mathbf{K}$, ϑ the angle between $\mathbf{K}$ and the x-axis,

$$K_x=K\cos\vartheta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.13)$$

and ϕ is an azimuth around the K-axis. Since $c_{\mathbf{w}}$ is independent of this azimuth ϕ we can average in (3.12) over ϕ, thus transforming $\mathbf{w}_x$ into $w\cos\theta\cos\vartheta$. With the use of (3.13) we then find

$$\frac{m}{\hbar}v_x-K_x=-\frac{K_x}{K}\sum_{\mathbf{w}} w|c_{\mathbf{w}}|^2\cos\theta\Big/\Big(1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2\Big). \quad . \quad . \quad (3.14)$$

On the other hand, from (2.27) we find by using (2.26),

$$\frac{m}{\hbar^2}\frac{\partial}{\partial K_x}(E(K)-E_0(K))-\frac{m}{\hbar^2}\frac{\partial E(K)}{\partial K_x}-K_x=\frac{m}{\hbar^2}\frac{K_x}{K}\frac{\partial}{\partial K}\sum_{\mathbf{w}}\frac{|(e\Phi)_{\mathbf{w}}|^2}{E_0(\mathbf{K},\mathbf{w})-E(K)}$$
$$=\frac{K_x}{K}\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2\left(K-w\cos\theta-\frac{m}{\hbar^2}\frac{\partial E(K)}{\partial K}\right). \quad . \quad (3.15)$$

Hence the expression $(m/\hbar^2)(\partial E/\partial K_x)-K_x$ becomes equal to the right hand side of (3.14); equation (3.10) then follows immediately.

We shall now calculate the value of the normalization integral J (2.28) which is required to judge the accuracy of the perturbation method. By inserting from (2.26) with the use of (2.16), (2.11), (2.21) and (3.1) we can write

$$J=1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2=1+\frac{\alpha}{4\pi^2}u^3\int\frac{dw\,d\Omega}{[(\mathbf{K}-\mathbf{w})^2-K^2+u^2(1+y)]^2}. \quad . \quad (3.16)$$

A straight forward integration leads to

$$J=1+\frac{\alpha}{2(1+y)(1+y-x^2)^{\frac{1}{2}}}. \quad . \quad . \quad . \quad . \quad . \quad (3.17)$$

It will be noticed that at the lowest level $x=0$, $y=y_0$, the value of J becomes, using (3.5),

$$J_0=1+\frac{\alpha}{2(1+y_0)^{\frac{3}{2}}}=\frac{2+3y_0}{2(1+y_0)}. \quad . \quad . \quad . \quad . \quad (3.18)$$

Thus near $x=0$, the perturbation method converges if the interaction parameter α is sufficiently small compared with unity. As we approach the upper limit of our energy range, however, J tends towards infinity in view of (3.8). The perturbation method thus becomes increasingly unsatisfactory when the variational method should still be reliable. At the upper limit of our range $x=x_1$, however, J tends to infinity which means that here the variational method also breaks down.

A further quantity of interest is the expectation value of the polarization potential (2.4) at a certain distance from the electron. We have considered this for the lowest level, $K=0$, only. For this purpose we first require the quantity

$$\Phi(\mathbf{r},\mathbf{r}_e)=\int\Psi^*\Phi(\mathbf{r})\Psi\Big/\int\Psi^*\Psi, \quad . \quad . \quad . \quad . \quad (3.19)$$

where the integrals extend over all the coordinates except those of the electron. Thus with (2.4)

$$\Phi(\mathbf{r},\mathbf{r}_e)=\frac{4\pi}{V^{\frac{1}{2}}}\sum_{\mathbf{w}}\frac{1}{w}\left(X_{\mathbf{w}}(\mathbf{r}_e)\sin\mathbf{wr}+\frac{Y_{\mathbf{w}}(\mathbf{r}_e)}{M\omega}\cos\mathbf{wr}\right), \quad . \quad (3.20)$$

where we define

$$X_{\mathbf{w}}(\mathbf{r}_e)=\int\Psi^*X_{\mathbf{w}}\Psi\Big/\int\Psi^*\Psi, \quad . \quad . \quad . \quad . \quad . \quad (3.21)$$

and similarly for Y. Use of (2.22), (2.13) (replacing there $\mathbf{r}$ by $\mathbf{r}_e$) and (2.28) leads to

$$(1+\sum_{\mathbf{w}}|c_{\mathbf{w}}|^2)X_{\mathbf{w}}(\mathbf{r}_e)=(c_{\mathbf{w}}\exp\{-i\mathbf{w}\mathbf{r}_e\}+c^*_{\mathbf{w}}\exp\{i\mathbf{w}\mathbf{r}_e\})\left(\frac{\hbar}{2M\omega}\right)^{\frac{1}{2}}. \quad (3.22)$$

Here the square root is the ordinary oscillator matrix element for X, and the value (2.12) of M must be introduced. A similar expression is obtained for $Y_{\mathbf{w}}(\mathbf{r}_e)$. By inserting $c_{\mathbf{w}}$ from (2.26) the summation of (3.20) can be carried out with the help of (2.11). Making further use of (3.5) and (3.18) the result is

$$\Phi(\mathbf{r},\mathbf{r}_e)=-\left(\frac{1}{\epsilon_\alpha}-\frac{1}{\epsilon}\right)\frac{e}{|\mathbf{r}-\mathbf{r}_e|}(1-\exp\{-|\mathbf{r}-\mathbf{r}_e|u(1+y_0)^{\frac{1}{2}}\})\frac{2}{2+3y_0}. \quad . \quad . \quad . \quad . \quad (3.23)$$

We notice that the energy of interaction is given by

$$e\Phi(\mathbf{r}_e,\mathbf{r}_e)=-\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)e^2u\frac{2(1+y_0)^{\frac{1}{2}}}{2+3y_0}=-\frac{4(1+y_0)^{\frac{1}{2}}}{2+3y_0}W, \quad . \quad (3.24)$$

which for small values of y_0 becomes equal to $-2W$. A fraction of this energy has to be spent to polarize the lattice.

RESULTS AND DISCUSSION.

§ 4. PROPERTIES OF THE LOWEST ENERGY LEVELS.

The results which we have obtained in the previous calculations concern the lowest energy levels of the system consisting of one electron and of a continuous dielectric medium possessing a proper frequency $\omega/2\pi$. Apart from the electronic charge e and mass m (which might be replaced by an effective mass to account for the periodic lattice field) they contain three parameters all of which can be found experimentally. These are the static dielectric constant ϵ, with optical refractive index $\epsilon^{\frac{1}{2}}_\infty$ and the frequency $\omega/2\pi$ of long longitudinal polarization waves. The latter can be expressed by the frequency $\omega_t/2\pi$ of long transverse waves which is the experimentally known frequency of infra-red residual waves. One finds (Fröhlich 1949, § 18)

$$\omega=\omega_t(\epsilon/\epsilon_\infty)^{\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

Long wave in this connection means long compared with the lattice distance. The calculations have shown that the length b introduced for dimensional reasons in (1.4) is the characteristic length occurring in our results. The fact that its ratio to the lattice distance

$$\frac{b}{a}=\frac{1}{a}\left(\frac{\hbar}{2m\omega}\right)^{\frac{1}{2}}\geqslant 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.2)$$

is large compared with unity for any reasonable value of ω justifies our procedure to treat the material as continuous.

The method which we have adopted contained a development of the polarization of the dielectric in normal coordinates. In the absence of electrons the behaviour of a dielectric material is then described in terms of quantized polarization waves. All quanta have the same energy $\hbar\omega$, but their momenta may differ, depending on the wavelength. The interaction with an electron can then be described in terms of absorption and emission of quanta $\hbar\omega$ by the electron; this is in fact the procedure previously used to calculate the scattering of electrons (*cf.* I.). At present we are however, mainly concerned with the energy levels of our system. Their broadening by scattering processes would lead us at once into difficulties. We have escaped this, in the present paper, by restricting ourselves to the lowest energy levels in a range $\hbar\omega$ above the ground level; in this case a free quantum can never exist. It should be realized that energy levels always refer to the energy of the whole system. To assume that not a single free quantum is excited in an infinite dielectric can only mean that its temperature is zero. We shall find below, however, that actually the only restriction required is to assume that

$$kT \ll \hbar\omega. \qquad \qquad (4.2a)$$

In the mathematical technique which we have adopted we chose solutions which did not allow the existence of free quanta; for technical reasons they break down at an energy somewhat higher than $\hbar\omega$ above the ground level. We shall present them in this whole range, but it should always be kept in mind that they are reliable only if the energy difference ΔE from the ground level is smaller than $\hbar\omega$,

$$\Delta E < \hbar\omega. \qquad \qquad (4.3)$$

Lowest Levels.—According to (3.5), using (3.1) and (2.19) the lowest energy level $E(0)$ is obtained from the equation

$$-E(0)[1-E(0)/\hbar\omega]^{\frac{1}{2}} = W, \qquad \qquad (4.4)$$

i. e.

$$E(0) \simeq -W \quad \text{if} \quad W \ll \hbar\omega, \qquad \qquad (4.5)$$

and

$$E(0) \simeq -(\hbar\omega)^{\frac{1}{3}} W^{\frac{2}{3}} \quad \text{if} \quad W \gg \hbar\omega. \qquad \qquad (4.6)$$

The energy W depends on the three parameters ϵ, ϵ_∞ and ω according to equations (1.5) and (1.4). For practically interesting cases usually $\alpha = W/\hbar\omega$ is somewhat larger than unity (*cf.* §5.) Also it should be remembered that $W \propto \omega^{\frac{1}{2}}$, *i. e.* $\alpha \propto 1/\omega^{\frac{1}{2}}$.

The lowest level is the lower limit of a continuum; the energy difference from the lowest level, $\Delta E(K)$, depends on the parameter K when $\mathbf{K}$ is the momentum of our system. This should be distinguished from the electronic contribution $m\mathbf{v}$($\mathbf{v}$=electronic velocity) because the dielectric carries a part of the momentum. It follows from (3.1) and (2.19) that

$$\Delta E(K) = E(K) - E(0) = \frac{\hbar^2}{2m} K^2 + \hbar\omega(y_0 - y), \qquad \qquad (4.7)$$

where y satisfies equation (3.4). In the neighbourhood of $K=0$, in particular, using (3.6) and (3.1),

$$\Delta E(K) \simeq \frac{\hbar^2}{2m} K^2 \left(1 - \frac{1}{3}\,\frac{y_0}{2+3y_0}\right). \qquad (4.8)$$

The second term in the bracket takes account of the influence of the interaction between electron and dielectric. Since the quantity $y_0=-E(0)/\hbar\omega$ is positive the value of the bracket is always between 1 and 8/9. Thus, though the shift of an energy level due to interaction may be considerable (compared with $\hbar\omega$) it depends only very little on the momentum.

It should be mentioned here that ordinary second order perturbation gives the value (4.5) for the lowest level, and replaces the above bracket by $(1-y_0/6)$. If this method were applied when the condition $y_0<1$ does not hold it would lead to the contradictory result that the state $K=0$ is not the lowest. This demonstrates the supremacy of the variational method.

Turning to larger values of K, the deviations of $\Delta E(K)$ from $\hbar^2K^2/2m$ become increasingly larger. In fact the curvature of the curve representing $\Delta E(K)$ against K changes sign and ends with horizontal tangent at a value $K_1=ux_1$ (*cf.* 3.1) of K which constitutes the limit beyond which the mathematical formalism breaks down. At this value, according to (3.8) and (3.1),

$$E(K_1)=\hbar\omega. \qquad (4.9)$$

Since $E(0)<0$ it follows that always

$$\Delta E(K_1)>\hbar\omega. \qquad (4.10)$$

Hence (*cf.* 4.3) this upper region is outside the range of validity of our calculations.

Velocity.—Making use of (3.10) and (4.8), the average electronic velocity near the ground level is given by

$$v=\frac{\hbar K}{m}\left(1-\frac{1}{3}\,\frac{y_0}{2+3y_0}\right). \qquad (4.11)$$

For higher values it reaches a maximum and then falls to zero. Again these latter parts are outside the range of validity of our calculations.

Acceleration. An external field F accelerates our system. To discuss this question we imagine a wave packet to be formed so that we can approximately fix the position of the electron. Shifting it a distance Δx means a transfer of energy $eF\Delta x$ from the field to the electron. Hence, the value of K_x changes by an amount ΔK_x and the energy-treating ΔK_x as small—by $\Delta K_x \partial E(K)/\partial K_x$. Dividing by the time Δt required, we have with $v_x=\Delta x/\Delta t$ and with (3.10),

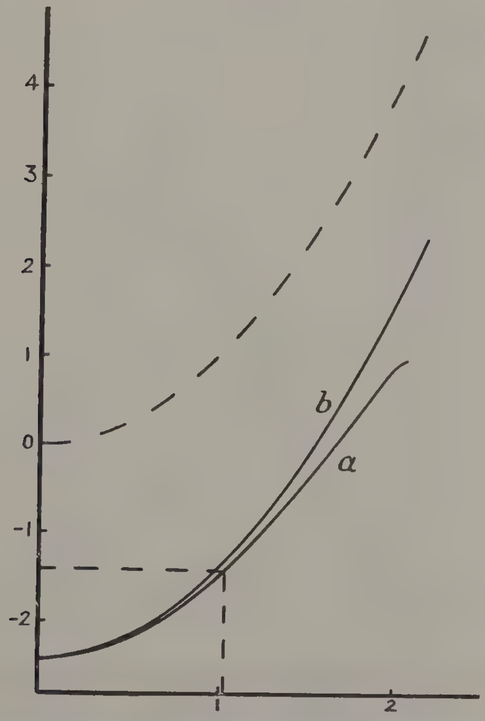
$$\frac{\Delta K_x}{\Delta t}=\frac{eF}{\hbar}, \qquad (4.12)$$

an equation well known from the theory of metals. Using again (3.10) we find

$$\frac{\Delta v_x}{\Delta t}=\frac{1}{\hbar}\frac{\partial^2 E}{\partial K_x^2}\frac{\Delta K_x}{\Delta t}=\frac{eF}{\hbar^2}\frac{\partial^2 E}{\partial K_x^2}. \quad . \quad . \quad . \quad . \quad (4.13)$$

In the region in which (4.8) and (4.11) hold the acceleration is thus similar to that of a free electron. For larger values of K it decreases, reaches zero when the velocity has its maximum and then reverses its sign. The latter region again is outside the range of validity of our calculations.

Fig. 1.

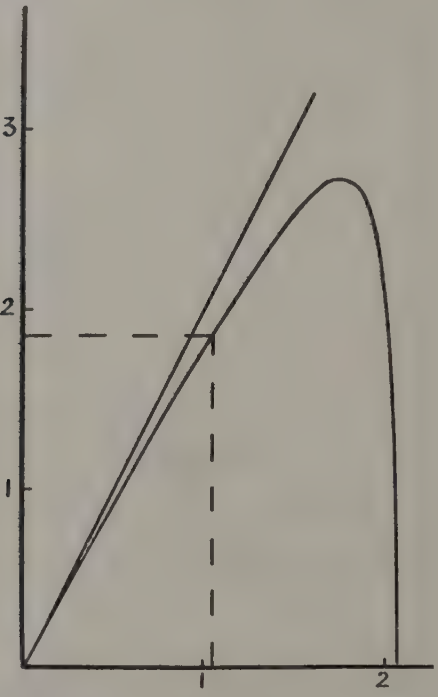


Curve a represents the energy of electrons in units of $\hbar\omega$ (*i. e.* the quantity x^2-y) as a function of the reduced momentum x; the dotted curve refers to free electrons. Curve b is the same curve shifted so as to demonstrate that the interaction hardly depends on the momentum. Validity outside the dotted region is doubtful.

The main results of our calculations are shown in the accompanying figures. The quantities $\hbar\omega$ and $\hbar u$, respectively, have been used as units of energy amd momentum. The results according to §3 then depend on the interaction parameter only. The value $\alpha=4{\cdot}47$ has been used in the drawings (for NaCl it would be 5·2). Fig. 1 represents the energy as function of the momentum For comparison a free electron curve has been drawn showing a shift considerably larger than $\hbar\omega$ (*i. e.* unity in the present units). A displaced free electron curve starting at the same ground level demonstrates that the shift due to interaction is nearly independent of the momentum. The validity of the energy curve is restricted to values in a range $\hbar\omega$ above the ground level.

Fig. 2 shows the average electronic velocity as a function of momentum, compared with the case of free electrons. Within the range of validity the deviations are seen to be small.

Fig. 2.



The full line is the average electronic velocity as a function of momentum; the straight line corresponds to free electrons. Validity outside dotted region is doubtful.

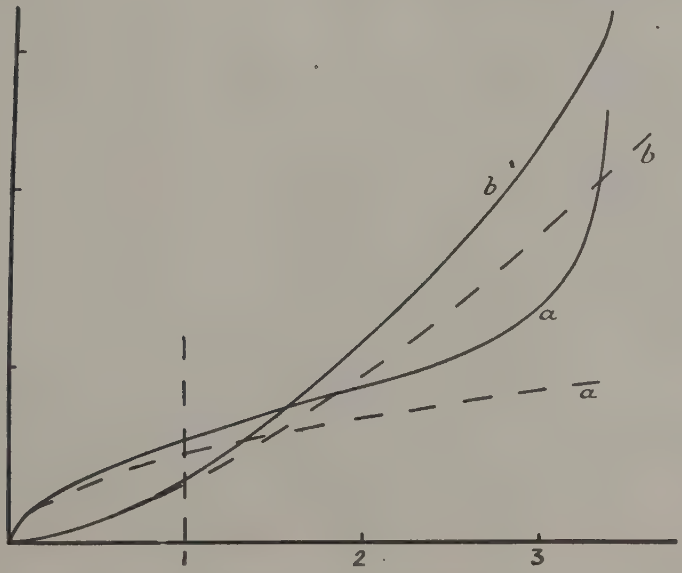
In fig. 3 we show the density of energy levels per unit energy $D(E)$ as a function of energy, and the total number of levels up to a given energy, $\int D(E)\,dE$. For comparison the corresponding curves in the case of free electrons are also drawn. The quantity $D(E)$ can easily be calculated from relation

$$D(E) \propto K^2 \frac{dK}{dE} \cdot \qquad (4.14)$$

§ 5. The Question of Self-trapping.

As mentioned at the end of § 1 it has first been suggested by Landau (1933) that an electron might "dig its own hole" and thus be trapped in the medium. The electron would thus be described by a probability distribution whose value is appreciable only in the neighbourhood of a given point in the medium. It will then give rise to a certain polarization which if frozen in, would act as a centre of attraction for the electron and lead to at least one bound state. Probability distribution and polarization can be made self consistent. The result is a particular solution of (2.6).

Fig. 3.



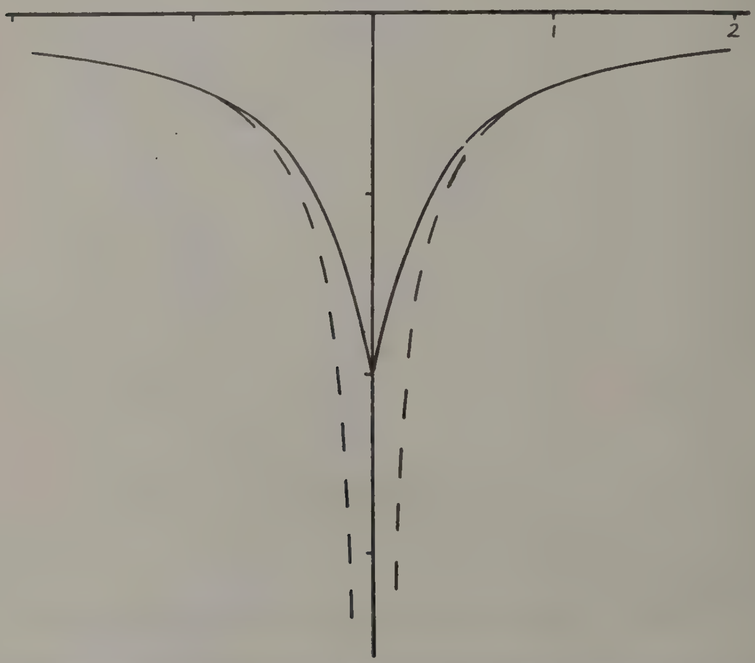
Curve *a* represents the density of energy levels as a function of energy ; *b* shows the total number of levels. The dotted curves refer to free electrons. Validity on the right of the vertical line is doubtful.

One of the objections to be raised against the above method is the preference given to a particular point in the medium. No doubt, in the exact solutions for stationary states the electron can be found with equal probability at any point. Our solution (2.22), (2.26) satisfies this requirement. At the same time the polarization relative to the position $\mathbf{r}_e$ of the electron (*cf.* fig. 4) shows a very similar behaviour as it does in the Landau method relative to a fixed point. The polarization potential $\Phi(\mathbf{r}, \mathbf{r}_e)$ behaves for large distances like $1/|\mathbf{r}-\mathbf{r}_e|$ down to distances of the order b (provided α is not very big). For smaller distances $\Phi(\mathbf{r}, \mathbf{r}_e)$ varies only slowly as compared with $1/|\mathbf{r}-\mathbf{r}_e|$.

A second objection arises when the linear extension of the probability distribution is larger than our characteristic distance b, equation (1.4). In this case it is no longer possible to assume the polarization to be rigid. This objection mainly holds for small values of the interaction parameter α [*cf.* (3.1)], or if α is large, for excited states whose orbit is larger than b.

Comparison of our results for the lowest energy level $E(0)$ [equation (4.4)] with those obtained by Pekar (1946) (using the Landau method) confirms the above conjecture in the case that the quantity $c=(1/\epsilon_\infty-1/\epsilon)^{-1}$, and hence (*cf.* 3.1) α, is small compared with unity. For in this case Pekar's value for $|E(0)|$ is proportional to c^2 and therefore of a smaller order than ours which [*cf.* (4.5) and (1.5)] is proportional to the first power of c. Our value for $|E(0)|$ on the other hand, being based on a variational method, is certainly not too large.

Fig. 4.



The full curve represents the average potential (3.23) as a function of a reduced distance $\xi=|\mathbf{r}-\mathbf{r}_e|u(1+y_0)^{\frac{1}{2}}$ from the electron; the dotted curve refers to the case of a point charge.

For strong interaction (large α), both methods lead to similar values for $E(0)$. Thus Markham and Seitz (1948) using methods developed by Mott and Littleton (1938) and by Pekar (1946) find the value of $-0{\cdot}13$ eV. as compared with $-0{\cdot}09$ in our calculations*.

---

* Obtained from equation (4.4) with the help of (1.5) and (1.4) using the following numerical values: $\epsilon=5{\cdot}6$, $\epsilon_\infty=2{\cdot}3$, $\omega=4{\cdot}8\times10^{13}$ sec^{-1}. The latter [*cf.* (4.1)] corresponds to a residual wavelength $(2\pi c_0/\omega)(\epsilon/\epsilon_\infty)^{\frac{1}{2}}=61\mu$, c_0=velocity of light.

The case of the Landau method for a calculation of the lowest energy level thus can be expected to lead to reasonable values only if the interaction parameter α is large. This method as described so far does not yet lead to self-trapping. Two further assumptions would have to be made which have never been proved. They are both objectionable and our calculations prove in fact that they are wrong. They are

(i.) that the whole configuration cannot move freely through the lattice,
(ii.) that there would be a continuum above the (presumed) discrete ground level similar to the case of an electron moving in a permanent field of force.

This latter assumption would lead to an infinitely extended probability distribution for which, as we have shown above, the Landau method cannot be applied.

Assumption (i.) is disproved by our calculations which show that the energy difference of states with non-vanishing momentum $\hbar K$ from the ground state ($K=0$) hardly differs from that for free electrons although the shift of energy levels due to the interaction may be large. That assumption (i.) cannot be correct can also be seen more directly from the fact that the polarization due to a moving point charge differs from that of a static one only at distances below v/ω. For an extended charge distribution, representing the electron in the Landau method, no appreciable dependence of the polarization on the velocity can be expected so long as v/ω is smaller than the main extension of the charge. In this case, then, the interaction energy (including the energy required to polarize the medium) is nearly independent of velocity.

It should be added that our method appears to be quantitatively correct for values of α up to $\alpha \simeq 1$. For larger values ($\alpha \simeq 5$ for NaCl) we hope that the variational method still ensures a correct order of magnitude. Improved calculations are in progress to investigate this.

§6. MEAN FREE PATH.

In calculating the conductivity the average contribution per electron is usually written in the form $e^2\bar{\tau}/m$ where $\bar{\tau}$ is a quantity with the dimension of a time, the average relaxation time. The quantity $e\bar{\tau}/m$ is often denoted as mobility. If a model is used in which electrons are free to move except for occasional elastic collisions, then a relaxation time $\tau(v)$, depending on the velocity v can be introduced in such a way that its average over all electrons, $\overline{\tau(v)}$ is equal to $\bar{\tau}$,

$$\overline{\tau(v)}=\bar{\tau}. \qquad \text{(6.1)}$$

Similarly a mean free path $l(v)$ and an average mean free path $\bar{l}$ can be introduced by

$$l(v)=\tau(v)v, \qquad \bar{l}=\overline{l(v)}. \qquad \text{(6.2)}$$

The relaxation time $\tau(v)$ is not simply the average time between two collisions but contains also the scattering angle ; it might be called the average time between two large angle scatterings.

In our case of electrons in an ionic crystal the assumption that the interaction between electrons and polarization waves can be treated as a small perturbation leads to a straight forward calculation of the conductivity, and hence of the relaxation time although the collisions are not elastic and are connected with absorption and emission of quanta $\hbar\omega$. It turns out, however, that only for electronic energies large compared with $\hbar\omega$ (treated in I.) or smaller than $\hbar\omega$ (Fröhlich and Mott 1939) is this treatment reliable. Energies which are only slightly larger than $\hbar\omega$ lead to relaxation times which are smaller than $1/\omega$ or to mean free paths shorter than the de Broglie wavelengths as we shall see presently. This means that the method is then no longer valid.

To treat this case of energies of several $\hbar\omega$ the calculation in (I.) must be slightly modified. For in (I.) we were mainly concerned with fast electrons scattered primarily by very short polarization waves. Electrons with energies of only several $\hbar\omega$, however, absorb or emit only very long polarization waves as follows from conservation of energy and momentum This means that the assumptions made at the beginning of this paper (especially §2) hold. Instead of carrying through the calculations in detail it is sufficient to compare with (I.) the expressions for the matrix elements and from there derive the replacements to be made in the final expression. The reason is that the Hamiltonian used in (I.) differs from ours only in the value of constant parameters. Comparing now the expression for $M_{kk'w}$ in (I. page 236) with our matrix element $(e\Phi)_{\mathbf{KK'w}}$, [equations (2.15) and (2.16)] we see that we have to replace

$$(a^3N)^{\frac{1}{2}}, \qquad \frac{e^2}{a^{\frac{3}{2}}}, \qquad \frac{1}{(M\nu)^{\frac{1}{2}}},$$

of (I.) by

$$V, \qquad 2^{\frac{1}{2}}e, \qquad \left[\frac{\omega}{2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\right]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (6.3)$$

respectively. Thus equations (15) and (16 *a*) of (I.) must be replaced by

$$\frac{1}{\tau}=\frac{1}{\tau_0}\left[1+\frac{2}{\exp\{\hbar\omega/kT\}-1}\right] \quad . \quad . \quad . \quad . \quad . \quad (6.4)$$

and

$$\frac{1}{\tau_0}=\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\frac{e^2\omega}{\hbar v}, \quad \text{if} \quad E'>m\frac{v^2}{2}\gtrsim \text{several } \hbar\omega. \quad . \quad . \quad (6.5)$$

The energy E' (denoted by $2^{\frac{2}{3}}\pi^2\hbar^2/(8ma^2)$ in I.) is of the order of 1 eV. In thermal equilibrium the condition attached to (6.5) holds for an appreciable number of electrons only if $kT \gg \hbar\omega$. It then follows that

$$\frac{1}{\tau_0}=\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\frac{2e^2kT}{\hbar^2 v}, \quad \text{if} \quad kT \gg \hbar\omega. \quad . \quad . \quad . \quad . \quad (6.6)$$

Hence using (6.2) and $\overline{v^2}=3kT/m$ we obtain

$$\bar{l}=\frac{3}{2}\frac{\epsilon\epsilon_\infty}{\epsilon-\epsilon_\infty}a_0 \quad \text{if} \quad kT \gg \hbar\omega, \quad . \quad . \quad . \quad . \quad . \quad (6.7)$$

where

$$a_0=\frac{\hbar^2}{me^2}\simeq 0{\cdot}5\times 10^{-8}\,\text{cm}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.8)$$

is the Bohr radius.

The remarkably simple result leads to values of $\bar{l}$ of the order of 5×10^{-8} cm. This is appreciably smaller than the de Broglie wavelength of thermal electrons at room temperature. The quantitative validity of (6.7) is, therefore, doubtful.

Electrons with energies below $\hbar\omega$ can only absorb but not emit quanta $\hbar\omega$. The perturbation calculation for this case has been carried through by Fröhlich and Mott (1939) modifying the method developed in (I.) so as to hold for very slow electrons. In the light of the present paper a further modification of the constant is required now, by making the replacement (6.3). This should be made in the second formula on page 501 of Fröhlich and Mott (1939) leading to replace

$$\frac{1}{\tau_0}=\frac{2}{3}\frac{e^4k}{Ma^3\hbar v\nu k_\nu} \quad \text{by} \quad \frac{1}{\tau_0}=\frac{2}{3}\frac{\omega}{a_0u}\left(\frac{1}{\epsilon_x}-\frac{1}{\epsilon}\right), \quad . \quad . \quad (6.9)$$

considering equation (6.8). Use has been made of the fact that k_ν of the previous paper has now been denoted by u, *cf.* equation (2.21), and that $\hbar k/m=v$. Applying then equation (9) of Fröhlich and Mott and (2.21) we obtain for the mean free path

$$l=\tau v=\tau_0 v(\exp\{\hbar\omega/kT\}-1)=3a_0\frac{\epsilon\epsilon_\infty}{\epsilon-\epsilon_\infty}\left(\frac{mv^2}{2\hbar\omega}\right)^{\frac{1}{2}}(\exp\{\hbar\omega/kT\}-1). \quad (6.10)$$

The average mean free path thus becomes

$$\bar{l}=\frac{6}{\sqrt{\pi}}a_0\frac{\epsilon\epsilon_\infty}{\epsilon-\epsilon_\infty}\left(\frac{kT}{\hbar\omega}\right)^{\frac{1}{2}}(\exp\{\hbar\omega/kT\}-1) \quad \text{if} \quad kT \ll \hbar\omega. \quad . \quad (6.11)$$

It is larger than the one obtained previously [equation (11)] by a factor

$$\frac{\epsilon\epsilon_\infty}{\epsilon-\epsilon_\infty+1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.12)$$

As a consequence of this correction there is a considerable improvement in the agreement between theoretical and experimental values of $\bar{l}$ of cuprous oxide. The previous formula gave values eight times too small; the factor (6.12) reduces this to about two ($\epsilon\simeq 10$). It is thus required to assume the effective mass to be only slightly smaller than the electron mass to obtain agreement.

The above formulæ for mean free paths have been derived on the assumption that perturbation theory is applicable. We found in §§ 2 and 3 that at the absolute zero of temperature this holds if the interaction parameter α given by equation (3.1) is small, $\alpha<1$. This result can easily be extended to higher temperatures as long as $kT \ll \hbar\omega$. For on this condition relatively few free quanta are excited; they influence only little the calculations of the energy of interaction so long as we deal with electrons whose kinetic energy, in zero approximation, is sufficiently far below $\hbar\omega$.

Thus although equation (6.11) seems well founded, if $\alpha < 1$, a serious objection can be raised. Absorption of a quantum $\hbar\omega$ leads the electron into a level from which after a very short interval it would re-emit a quantum $\hbar\omega$—in general with different momentum. This suggests to treat such collisions as second order transitions, leading to the absorption of a quantum $\hbar\omega$, followed by a competitive emission of another quantum $\hbar\omega$ whose wave number $\mathbf{w}$ can have any value in a whole range. The formal method to be adopted would be very similar to that used by Breit and Wigner (1936) in nuclear physics *.

It should also be clear that by allowing the temperature T to be different from zero the energy of our whole system is always considerably higher than $\hbar\omega$ above the ground level. We have just seen that for low temperatures this would not seriously affect our calculations as long as perturbation theory is valid. This holds also for larger values of the interaction parameter α where the variational method differs considerably from the perturbation method, but only as far as the energy of interaction is concerned. Collisions would require an entirely new treatment. The former contention follows at once by considering in the variational calculations, especially in the sum in (2.27), only those wave numbers for which free oscillations are not excited thermally and then treating the latter by perturbation theory. This is permissible because their relative number is small so long as $kT \ll \hbar\omega$; the subsequent change in the expression for the energy [*cf.* (3.3)] is therefore small too. In the case of collision, however, the method cannot be applied because it is then no longer possible to separate the treatment of excited and nonexcited oscillators.

The authors are indebted to the British Electrical and Allied Industries Research Association for their support.

REFERENCES.

BREIT and WIGNER, 1936, *Phys. Rev.*, **49**, 519.
FRÖHLICH, H., 1937, *Proc. Roy. Soc. A.*, **160**, 230; 1949, *Theory of Dielectrics*, Clarendon Press, Oxford.
FRÖHLICH H., and MOTT, N. F., 1939, *Proc. Roc. Soc. A*, **171**, 496.
FRÖHLICH, H., and PELZER, 1950, To be published; also E.R.A. Report, L/T 184.
LANDAU, 1933, *Z. Phys. Sowjet.*, **3**, 664.
MARKHAM and SEITZ, 1948, *Phys. Rev.*, **74**, 1014.
MOTT, N. F., and LITTLETON, 1938, *Trans. Faraday Soc.*, **34**, 485.
PEKAR, 1946, *Journ. Physics U.S.S.R.*, **10**, 341.

---

* *Added in proof.*—One of us (S.Z.) has now carried out this calculation. It leads to a further factor $\frac{1}{2}$ in equation (6.11).

XXI. *Kinetics of the Phase Transition in Superconductors.*

By A. B. Pippard,
Royal Society Mond Laboratory, Cambridge *.

[Received December 15, 1949]

Abstract.

A study is made of the influence of electromagnetic effects on the speed at which a transition can occur between the normal and superconducting states of a metal in a magnetic field, and it is concluded that these are powerful enough to be the dominant factor determining the speed. Illustrative examples include the transition of a superconducting plane slab and cylinder in a field greater than critical, the destruction of superconductivity in a wire by means of a current, and the mechanism whereby the intermediate state is established.

I. Introduction.

The phase-transition of a superconductor into the normal state through the influence of a magnetic field has been extensively studied from the point of view of determining the conditions under which the two phases may exist side by side in equilibrium. Typical of the experimental work on this problem are determinations of critical magnetic fields; studies of the destruction of superconductivity in thin wires by passage of a current; and investigations of the intermediate state in bodies of non-zero demagnetizing coefficient. At the same time fundamental thermodynamical studies of the various problems have served to correlate the results of these experiments with other properties of the superconductor, such as the specific heat. Very much more attention has been directed to this, the statical aspect of the problem, than to the kinetic aspect, and scant consideration has been given to the mechanisms which are involved during the progress of the transition. Examples of the types of phenomena which demand kinetic treatment are such problems as the rate at which the transition from the superconducting to the normal state occurs in a field greater than the critical field, or from the normal to the superconducting state in a field less than critical; the closely allied problem of the rate of propagation of a phase boundary along a wire; and the problem of the mechanism whereby the intermediate state is established. Of the experimental work on this aspect of the phase transition we may cite as examples the investigation by Lazarev, Galkin and Khotkevich (1947) of the behaviour of wires carrying high-frequency alternating currents of such amplitude as to induce a phase transition, and the recent work of

---

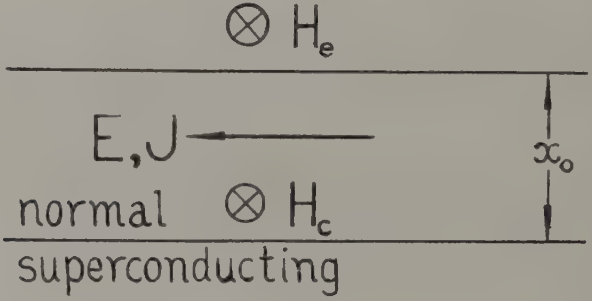
* Communicated by the Author.

Faber (1949) on the propagation of phase boundaries along wires. There appears to have been no discussion given of the electromagnetic processes which will have to be considered in any attempt to account for the observed phenomena, and the present paper represents an attempt to provide a starting-point for more rigorous treatments than can be given at the moment. It is unfortunate that the two experiments quoted are concerned with transitions which will be difficult to treat rigorously, and it is therefore not considered expedient to give any discussion of them here. On the other hand, the physical principles of the electromagnetic processes underlying these complicated examples may be understood clearly from idealized simple models, and the way may therefore be paved towards at least a qualitative understanding of the problems of more practical interest.

II. The Phase Transition in a Plane Semi-infinite Slab.

Consider a semi-infinite slab of a superconducting material having a plane surface normal to the X-axis, as in fig. 1. If a magnetic field H_e, greater than the critical field H_c, be suddenly applied parallel to the surface, we may imagine that a thin layer at the surface makes a transition into the normal state, so that there is created a plane phase boundary parallel to the surface. The boundary will move into the interior of the metal as the normal region expands, and we wish to find its rate of propagation. It is clear that so long as the external field exceeds H_c there can never be a stable position of the boundary, since if it comes to rest the field at the boundary will eventually rise to a value H_c. On the other hand, the rate of propagation cannot be too great, since as the normal region expands magnetic flux must enter it, and in doing so induce currents whose effect is to reduce the field strength at the boundary. It seems reasonable to suppose, then, that the boundary moves at such a rate that the field strength next to it is maintained through the action of the induced currents at a value H_c.

Fig. 1.



The fields in the normal metal must satisfy the equations (in which the units are electromagnetic):

$$\frac{\partial H}{\partial x}=4\pi J, \qquad \frac{\partial E}{\partial x}=\frac{\partial H}{\partial t}, \qquad \text{and} \qquad J=\sigma E.$$

Hence

$$\frac{\partial^2 H}{\partial x^2}=4\pi\sigma\frac{\partial H}{\partial t}.$$

If the field at the surface, H_e, be written as $H_c(1+p)$, and the field H at any point in the normal metal be written as $H_c(1+f)$, then f satisfies the same differential equation,

$$\frac{\partial^2 f}{\partial x^2}=4\pi\sigma\frac{\partial f}{\partial t}, \qquad \text{. . . . . (1)}$$

which must be solved subject to the following boundary conditions :--

(*a*) $f=p$ when $x=0$,

(*b*) $f=0$ when $x=x_0$, x_0 being the depth of the phase boundary below the surface of time t,

(*c*) $\frac{\partial f}{\partial x}=-4\pi\sigma\frac{dx_0}{dt}$ when $x=x_0$.

Condition (*c*) follows from the fact that the electric field in the normal metal next to the boundary is equal to $H_c\frac{dx_0}{dt}$. A solution * of equation (1) may be found in the form:

$$f=f(y), \qquad \text{where} \qquad y=x/x_0.$$

Substituting in equation (1), we find that

$$f''+2\pi\sigma y\frac{d}{dt}(x_0^2)\,.\,f'=0. \qquad \text{. . . . . (2)}$$

In order that f shall be a solution it is necessary that the coefficient of f' shall be time-independent. We may therefore write

$$\frac{d}{dt}(x_0^2)=\frac{\alpha p}{2\pi\sigma}, \qquad \text{where } \alpha \text{ is a constant;}$$

i. e.

$$x_0^2=\frac{\alpha pt}{2\pi\sigma}, \qquad \text{if } x_0=0 \text{ when } t=0. \qquad \text{. . . . . . (3)}$$

---

* After this solution had been derived, it was pointed out to me by Dr. P. M. Marcus that the problem is analogous to that of the freezing of a lake, which is discussed by Carslaw and Jaeger (1947). I feel that there are sufficient differences in the treatments to justify setting out the following method of solution in full.

The constant α may be determined from equation (2) and the boundary conditions. We have that

$$f''+\alpha p y f'=0.$$

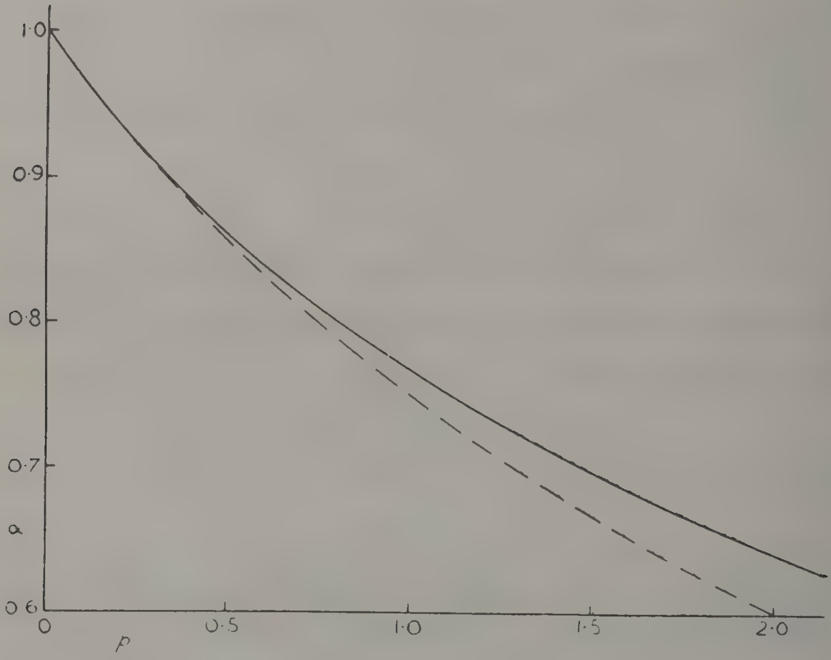
Hence, on integrating and applying condition (c),

$$f'=-\alpha p \exp\{\tfrac{1}{2}\alpha p(1-y^2)\}.$$

Integrating once more and applying conditions (a) and (b) we find that α must satisfy the equation

$$\alpha\int_0^1 \exp\{\tfrac{1}{2}\alpha p(1-y^2)\}\,dy=1.$$

Fig. 2.



This equation is readily solved numerically, and in fig. 2 the variation of α with p is exhibited. It will be seen that for small values of p, α may be taken as unity without great error, and equation (3), for the variation with time of the depth of the boundary below the surface, reads:

$$x_0^2 \doteqdot \frac{pt}{2\pi\sigma}. \qquad (4)$$

This approximate result might have been obtained very simply by supposing the field strength in the normal metal to be everywhere H_c.

For then a uniform current of density $\sigma H_c \frac{dx_0}{dt}$ will be induced by the movement of the boundary, so that in order to maintain a difference in field strength pH_c between the boundary and the surface it is necessary to put

$$4\pi\sigma H_c x_0 \frac{dx_0}{dt} = pH_c,$$

which leads immediately to equation (4). It will be found valuable to keep this approximate method in mind when we come to consider problems to which the above method of exact solution does not apply. The assumption that the field strength in the normal metal is everywhere H_c may be regarded as the first stage in the solution of the problem by successive approximation. We have seen that this assumption leads to a uniform current density in the normal metal, and we may now use this result to derive a closer approximation to the field distribution. The next stage in the calculation will thus start from the assumption that the field varies linearly from H_c at the phase boundary to $H_c(1+p)$ at the surface. A closer approximation to the current distribution may then be reached immediately, leading to the result of equation (3) with α put equal to $3/(3+p)$. This expression is shown as a dotted line in fig. 2, from which it is clear that the second approximation is close to the exact solution over a wide range of p.

It will be noted that the depth of the boundary, given by equation (3), varies as the square root of the time, as is to be expected from equation (1), which has the same form as the diffusion equation. Equations (3) and (4) bear a resemblance to the result obtained in the theory of the skin effect, that the depth of penetration of a field alternating with angular frequency ω is given by the expression :

$$\delta^2 = 1/2\pi\omega\sigma.$$

Thus, if we use the approximate solution (4), we see that for an external field $2H_c$ (*i. e.* $p=1$), the depth of the phase boundary after a time t is the same as the skin depth in the normal metal for a frequency $1/t$. For good conductors the rate of progress of the boundary is consequently very slow. For example, a specimen of pure tin may have in the normal state a conductivity of 5×10^8 ohm^{-1} cm.$^{-1}$, or in e.m.u., $\sigma = 0{\cdot}5$. If the external field exceeds the critical field by 10 per cent, $p = 0{\cdot}1$, and equation (3) becomes

$$x_0^2 \doteqdot 0{\cdot}03t.$$

A time of about 30 seconds is therefore needed for the boundary to penetrate 1 cm., and nearly one hour to penetrate 10 cm.

In this treatment we have considered the transition to be isothermal, although, as is well known, the adiabatic transition from the superconducting to the normal phase results in a drop in temperature. It is therefore necessary to consider whether it is justifiable to assume that if the outside surface of the material is maintained at constant temperature throughout, by contact with liquid helium, there is sufficiently good thermal conductivity to maintain isothermal conditions at all points. On account of the formal similarity between equation (1) and the equation of heat flow the analysis may be carried out by dimensional means. In order that the transition shall be isothermal, all that is necessary is that the skin depth for a heat wave at a given frequency shall be much greater than the corresponding skin depth for an electromagnetic field in the normal material. That is, $\left(\frac{4\pi\sigma\kappa}{\mathrm{C}}\right)^{\frac{1}{2}} \gg 1$, where κ is the thermal conductivity and C the specific heat per unit volume. For a good specimen of tin at 3° K the quantity on the left-hand side takes a value of about 500, so that the assumption of isothermal conditions is well satisfied. Since however $\sqrt{\sigma\kappa}$ is proportional to the conductivity, it might be necessary to consider thermal effects in specimens of poor conductivity.

III. The Phase Transition in a Cylinder.

Consider a long superconducting cylinder of radius a suddenly subjected to a longitudinal field $\mathrm{H}_e(=\mathrm{H}_c(1+p))$. Ideally we may suppose that a normal sheath will be formed at the surface, and the cylindrical superconducting core will shrink until finally it vanishes at the centre. The equation governing the rate of shrinkage will be the same as equation (1) but now on account of the cylindrical symmetry the method of solution adopted for the plane slab becomes inapplicable. We shall therefore adopt the approximate method mentioned in the last section to solve the problem, assuming the field strength in the normal material to take everywhere the value H_c.

Let the radius of the superconducting core be r_0; then the magnetic flux contained in the normal material between the core and a cylinder of radius r:

$$\phi=\pi(r^2-r_0^2)\mathrm{H}_c.$$

As the core contracts, the induced electric field at a radius r will be given by

$$\mathrm{E}(r)=-\frac{1}{2\pi r}\frac{d\phi}{dt}=\mathrm{H}_c\frac{r_0}{r}\frac{dr_0}{dt},$$

The circulating current density at radius r is thus given by

$$\mathrm{J}(r)=\sigma\mathrm{H}_c\frac{r_0}{r}\frac{dr_0}{dt}.$$

Now if the external field is $\mathrm{H}_c(1+p)$ we must have that

$$4\pi\int_{r_0}^{a}\mathrm{J}\,dr=-p\mathrm{H}_c.$$

Hence

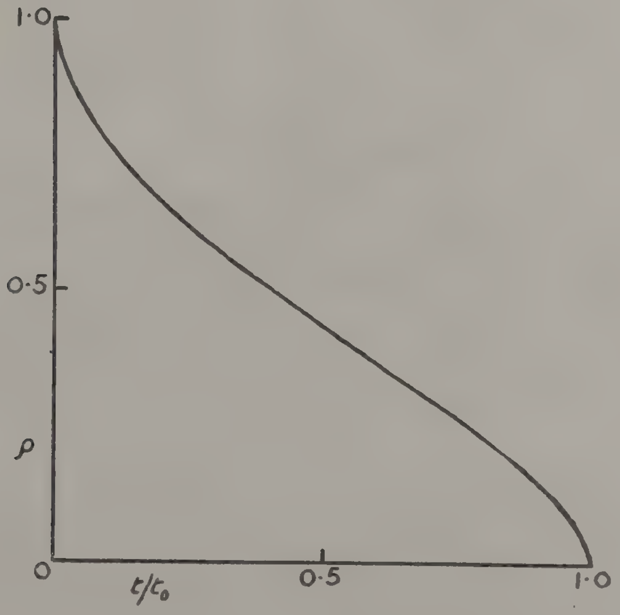
$$r_0\log\frac{a}{r_0}\cdot\frac{dr_0}{dt}=-p/4\pi\sigma.$$

Integrating, and putting $r_0=a$ when $t=0$, we obtain the result :

$$1-\rho^2(1-2\log\rho)=\frac{t}{t_0}, \qquad \text{where} \qquad \rho=r_0/a \qquad \text{and} \qquad t_0=\pi\sigma a^2/p.$$

When $\rho=0$, $t=t_0$, so that t_0 represents the total time required for destruction of superconductivity in the cylinder. Comparing this result with equation (4), we see that it is just one-half of the time needed for the penetration of a plane boundary to a depth r_0. The manner in which the superconducting core shrinks is illustrated in fig. 3, where ρ is plotted against t/t_0. The initial stage of the shrinkage approximates closely to the behaviour of the plane boundary, but after the radius of the core has fallen below one-half of the specimen radius the rate of shrinkage increases one more, until in the final stage the last traces of superconductivity vanish rapidly.

Fig. 3.



The reverse process, the transition from the normal to the superconducting phase in a field less than critical, is complicated by the occurrence of " supercooling ", the metastable persistence of the normal phase in a field slightly less than critical. It is no part of this discussion to enquire into the mechanism responsible for supercooling, which can only be understood in terms of a detailed theory of the superconducting state, and of the surface energies at the boundaries between the two phases.

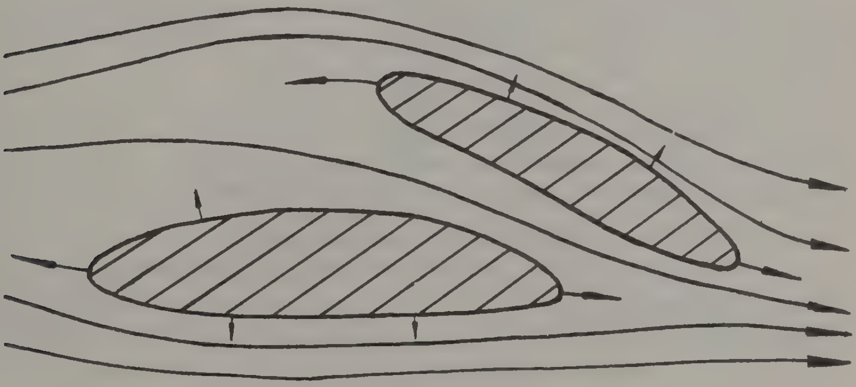
If, however, we suppose that eventually the supercooled phase breaks down by the formation of a small superconducting nucleus, its subsequent growth is probably largely determined by electromagnetic forces, and may be dealt with by the methods already described. The shape of the curve of fig. 3 suggests that the initial stage of growth may be extremely rapid, but we must remember that the figure refers to a long cylinder rather than to a small nucleus of approximately spherical shape. The nucleus will suffer from demagnetizing effects which will tend to increase the field strength at the magnetic equator, and decrease it at the poles. Thus the tendency will be for the most rapid expansion to occur along the magnetic lines of force, so that a long superconducting spindle will be formed, which will then expand, if it lies at the centre of the cylinder, after the manner of fig. 3 (the direction of the time axis being of course reversed).

IV. THE ESTABLISHMENT OF THE INTERMEDIATE STATE.

So far we have considered only examples of the phase transition in a specimen of zero demagnetizing coefficient, for which ideally the complete transition should be possible at the field strength H_c. In bodies of more complicated shape, such as ellipsoids, the transition in an increasing field can occur reversibly only if it is spread out over a range of field strengths, and in this range, as has been shown directly by Meshkovsky and Shalnikov (1947), the material is split into alternate regions of superconducting and normal material, a condition known as the Intermediate State. A theory of the equilibrium configuration of the intermediate state has been given by Landau (1943), and in this section we shall attempt a qualitative explanation of the mechanism whereby it is established. Let us consider a plane parallel slab of normal material in a magnetic field whose direction is perpendicular to the surface of the slab. As the field strength is reduced below H_c there will be an initial stage during which nothing occurs, the material remaining in the supercooled normal state. Eventually, however, we may expect a nucleus of superconducting material to be formed, which will, in accordance with the ideas of the previous section, expand rapidly along the field lines, and not so rapidly in a radial direction, until at last a spindle will extend right aross the specimen. This spindle should continue to expand radially until it takes up such a shape that over the whole interface between the normal and superconducting regions (though not at the boundary of the specimen) the field strength is just equal to H_c. This represents a stable configuration of the spindle. Now, during the whole of this process, the magnetic field has been expelled from the normal region, so that the field strength is never reduced below the value which it had before the nucleus was formed. There seems to be no reason therefore why any more spindles should be formed in the material until a chance fluctation results in the creation of another nucleus. Nevertheless, it is observed in all studies of the intermediate state that once the supercooled phase has broken down there is a sudden transition into an intermediate state filling the whole of the specimen. It appears as though the production of one nucleus leads to a chain of nucleation processes, enabling superconducting regions to be formed through the entire specimen.

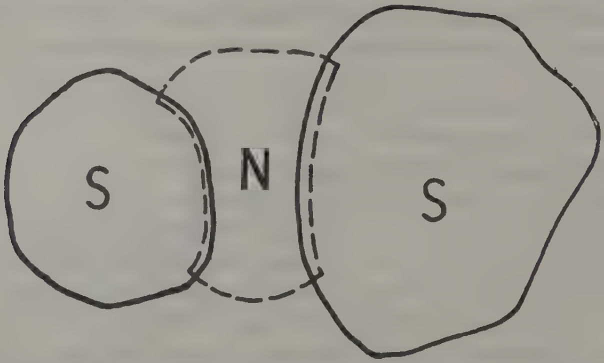
It is possible that the explanation of this effect is to be found in the imperfections of the specimen, such as grain boundaries or local strains, which lead to an irregular jerky growth of a superconducting spindle. If as the spindle grows along the magnetic lines it meets an impediment, so that its progress is momentarily retarded, demagnetizing effects will lead to a decrease of the field strength at its forward end. Just in front of the impeded spindle there will then be a region which is particularly favourable for the production of a new nucleus. Once a new nucleus has formed, further growth should proceed roughly as indicated in fig. 4, in which the arrows attached to the spindles represent the directions favourable to further growth. If any magnetic lines are trapped between the spindles, as will always happen unless the second nucleus is created exactly in front of the first spindle, there will arise electromagnetic forces which will effectively prevent coalescence of the two spindles. An adequate treatment of this problem, even by approximate methods, is difficult, but a qualitative argument should make the general principles clear. Consider the two spindles, S, shown in cross-section in fig. 5; for coalescence to occur it is necessary that the magnetic flux in the intervening region shall be ejected. Now the circuit shown dotted in the diagram may be regarded as an inductance, whose magnitude is determined by the area of the flux-containing region, N, of normal metal included in the circuit. As the spindles expand, L will vary approximately as the separation of the spindles. The resistance of the circuit will be determined by the length of the path in normal material between the spindles, and thus will also be roughly proportional to their separation. As the spindles expand, the time constant L/R for ejection of flux will therefore stay roughly constant, and the rate at which they approach one another will diminish continuously.

Fig. 4.



At the same time there is no such retardation of the growth in those directions in which there are no neighbouring spindles, so that coalescence is unlikely to occur before the spindles have expanded to such a size that the field strength in the normal environs has everywhere attained a value close to H_c, and further growth is impossible. We may therefore picture the establishment of the intermediate state as a spreading chain of nucleations, arising from the jerky growth of the initial and subsequent nuclei, until in the end the whole specimen is filled with superconducting spindles, and the intervening normal regions are permeated by a magnetic field of strength near H_c.

Fig. 5.



V. Destruction of Superconductivity by a Current.

As a last problem in the kinetics of phase boundary movement we shall consider the destruction of superconductivity in a thin cylindrical wire carrying a current. According to Silsbee's (1918) hypothesis, when the current reaches the value $\frac{1}{2}aH_c$, a being the radius of the wire, the field at the surface attains the critical value, and at this point resistance appears. Silsbee's hypothesis has been found by Scott (1948) to give a precise description of the critical current for destruction of superconductivity, even in very thin wires where the effects of surface energy between the phases might be expected to play some part. The behaviour of wires carrying currents greater than critical has been studied theoretically by F. London (1937), who has shown that over a certain range of currents an intermediate state must be set up. This is observed in practice, the resistance being only partially restored at the critical current, and then rising slowly to the normal value with increase of the current. There is not complete agreement between the experimental behaviour and London's prediction, and this is attributed by Scott to the effects of surface energy. This aspect of the problem will not concern us here, since we shall consider only the kinetic problem of the transition from the superconducting to the intermediate state, without any discussion of the exact nature of the intermediate state itself.

The problem differs from that of the cylinder in a longitudinal field, discussed in Section III, in that here we have a longitudinal current and a circulating magnetic field. We shall see that this profoundly modifies the results of the calculation. Let us at first neglect the possibility that during the movement of the phase boundary there is a return to the superconducting state of those regions of normal material which find themselves in a field less than H_c; that is, we shall suppose that at any instant during the transition the wire consists of an inner cylindrical core of superconducting material, of radius r_0, surrounded by a sheath of normal material. If the current is such as to produce a field at the surface of the wire of strength $H_c(1+p)$, then the total current must be

$$I_{tot}=\tfrac{1}{2}aH_c(1+p),$$

of which the part carried by the superconducting core

$$I_s=\tfrac{1}{2}r_0H_c.$$

in order to maintain a field strength H_c at the phase boundary. The remaining current will be carried in the normal metal,

$$I_n=\tfrac{1}{2}H_c\{a(1+p)-r_0\}. \qquad (5)$$

We shall now use the approximate method of solution which assumes the field strength in the normal metal to have everywhere the value H_c. It immediately follows that the electric field induced by a shrinkage of the core may be written in the form $E=-H_c\dfrac{dr_0}{dt}$, at all points in the normal metal. There is thus a uniform current density in the normal metal, $J_n=-\sigma H_c\dfrac{dr_0}{dt}$, and the total current carried by the normal metal,

$$I_n=-\pi\sigma H_c(a^2-r_0^2)\frac{dr_0}{dt}.$$

Comparing this result with equation (5) we have an expression for the rate of shrinkage,

$$-\frac{dr_0}{dt}=\frac{a(1+p)-r_0}{2\pi\sigma(a^2-r_0^2)}.$$

Integrating this equation, we find an expression for the time taken for the superconducting core to shrink to a radius r_0,

$$t/2\pi\sigma a^2=\tfrac{3}{2}+p-\rho(1+p+\tfrac{1}{2}\rho)+p(2+p)\log\frac{p}{1+p-\rho},$$

where

$$\rho=r_0/a. \qquad (6)$$

In particular, the time taken for the superconducting core to disappear entirely is given by

$$t_0/2\pi\sigma a^2=\tfrac{3}{2}+p+p(2+p)\log\frac{p}{1+p}.$$

The detailed form of this function is of little validity, as we shall see when we come to discuss the justification for the approximate method of solution, but its interest lies in the fact that if we put p equal to zero, we find a finite value, $3\pi\sigma a^2$, for t_0. Thus the calculation predicts that, in contrast to the behaviour of a cylinder in a longitudinal field, as soon as the current reaches the critical value, the transition may be completed in a finite time. For wires of the size used in experiments on the destruction of superconductivity by a current this time is very short. If the radius of the wire is ½ mm., and the material pure tin, the transition should be completed in 12 milliseconds.

So far it has been assumed that the transition consists solely of the shrinkage of the superconducting core until it finally disappears, but this cannot be the complete solution to the problem, since London has shown that stability is only achieved by the formation of an intermediate state. In order to understand how the intermediate state is set up, we must look into the magnetic field distribution in the normal metal. The field strength has been taken as H_c everywhere, but this assumption is not justified. In fact, if we now use our result that the current density is uniform in the normal metal to calculate the magnetic field, we find that when $p=0$ the field varies with radius according to the law

$$H(r)=H_c\frac{ar_0+r^2}{r(a+r_0)}. \qquad \qquad (7)$$

Thus H takes the value H_c when $r=r_0$ or a, in accordance with the boundary conditions, but is less than H_c at all intermediate radii. The minimum value of H occurs at a radius $\sqrt{ar_0}$, and at this point

$$H_{\min}=H_c\frac{2\sqrt{ar_0}}{a+r_0}.$$

As the core contracts, then, there are left behind regions of normal metal where the field strength is less than H_c, and when $r_0 \ll a$, there are regions near the centre of the wire where the field is very low. Presumably it is here that nucleation may occur for the establishment of the intermediate state, but no attempt has been made to carry the argument further, as it is difficult to conjecture the configuration of normal and superconducting regions which is stable in a wire carrying a current.

The manner in which the magnetic field is decreased below H_c in the normal metal suggests that the currents induced by the shrinkage of the core may have been overestimated, and hence the rate of shrinkage underestimated. A second approximation may be calculated by starting from the assumption that the field in the normal metal is given by equation (7), and it is found that in the final stage of collapse of the core the rate is indeed greater than was found in the first approximation. But the calculation is laborious, and the final result is complicated and mathematically not very well-behaved, so that is seems hardly worth while attempting any further stages of approximation.

The physical reason for the difference in behaviour of a cylinder when carrying a current and when subjected to a longitudinal field may be seen by considering the stability of the system under critical conditions. In the latter case, when the field strength is H_c, any diameter of superconducting material is stable, and the system may be said to be in neutral equilibrium. On the other hand, if we pass a critical current through the cylinder, and then suppose a normal sheath to be formed, eventually all the current will redistribute itself on the superconducting core of smaller diameter, producing there a field greater than H_c. A wire carrying the critical current is thus in unstable equilibrium, and as soon as any portion of the surface becomes normal there is a tendency for the surface current density in that region to increase and precipitate the transition into the normal state. This is probably the reason why Silsbee's hypothesis applies very well even in extremely thin wires, of the order of diameter of those which show effects of surface energy in transverse magnetic fields. The surface energy is unlikely to be great enough to counteract the instability of a wire carrying the critical current, though, of course, it will play a part in determining the configuration of normal and superconducting regions in the intermediate state, and hence the way in which the resistance approaches the normal value once the critical current has been exceeded. Such an hypothesis would be consistent with the observations of Scott.

I should like to thank Mr. T. E. Faber and Dr. B. Serin for interesting discussions on some of the points raised in this paper.

REFERENCES.

CARSLAW, H. S. and JAEGER, J. C., 1947, *Conduction of Heat in Solids*, p. 71, (Clarendon Press, Oxford).
FABER, T. E., 1949, *Nature*, **164**, 277.
LANDAU, L., 1943, *Jour. of Physics* (USSR), **7**, 99.
LAZAREV, B. G., GALKIN, A. A., and KHOTKEVICH, V. I., 1947, *C.R. Acad. Sci.*, USSR., **55**, 805.
LONDON, F., 1937, *Une conception nouvelle de la supraconductibilité* (Hermann et Cie.)
MESHKOVSKY, A. and SKALNIKOV, A., 1947, *Jour. of Exp. and Theor. Physics* (USSR), **17**, 851.
SCOTT, R. B., 1948, *Bull. Nat. Bur. of Standards*, **41**, 581.
SILSBEE, F. B., 1918, *Bull. Nat. Bur. of Standards*, **14**, 301.

XXII. *Programming a Computer for Playing Chess* *.

By Claude E. Shannon,
Bell Telephone Laboratories, Inc., Murray Hill, N.J. †

[Received November 8, 1949.]

1. Introduction.

This paper is concerned with the problem of constructing a computing routine or " program " for a modern general purpose computer which will enable it to play chess. Although perhaps of no practical importance, the question is of theoretical interest, and it is hoped that a satisfactory solution of this problem will act as a wedge in attacking other problems of a similar nature and of greater significance. Some possibilities in this direction are :—

(1) Machines for designing filters, equalizers, etc.

(2) Machines for designing relay and switching circuits.

(3) Machines which will handle routing of telephone calls based on the individual circumstances rather than by fixed patterns.

(4) Machines for performing symbolic (non-numerical) mathematical operations.

(5) Machines capable of translating from one language to another.

(6) Machines for making strategic decisions in simplified military operations.

(7) Machines capable of orchestrating a melody.

(8) Machines capable of logical deduction.

It is believed that all of these and many other devices of a similar nature are possible developments in the immediate future. The techniques developed for modern electronic and relay type computers make them not only theoretical possibilities, but in several cases worthy of serious consideration from the economic point of view.

Machines of this general type are an extension over the ordinary use of numerical computers in several ways. First, the entities dealt with are not primarily numbers, but rather chess positions, circuits, mathematical expressions, words, etc. Second, the proper procedure involves general principles, something of the nature of judgment, and considerable trial and error, rather than a strict, unalterable computing process. Finally, the solutions of these problems are not merely right or wrong but have a continuous range of " quality " from the best down to the worst. We might be satisfied with a machine that designed good filters even though they were not always the best possible.

---

* First presented at the National IRE Convention, March 9, 1949, New York, U.S.A.

† Communicated by the Author.

The chess machine is an ideal one to start with, since: (1) the problem is sharply defined both in allowed operations (the moves) and in the ultimate goal (checkmate); (2) it is neither so simple as to be trivial nor too difficult for satisfactory solution; (3) chess is generally considered to require " thinking " for skilful play; a solution of this problem will force us either to admit the possibility of mechanized thinking or to further restrict our concept of " thinking "; (4) the discrete structure of chess fits well into the digital nature of modern computers.

There is already a considerable literature on the subject of chess-playing machines. During the late 18th and early 19th centuries, the Maelzel Chess Automaton, a device invented by von Kempelen, was exhibited widely as a chess-playing machine. A number of papers appeared at the time, including an analytical essay by Edgar Allan Poe (entitled Maelzel's Chess Player) purporting to explain its operation. Most of these writers concluded, quite correctly, that the Automaton was operated by a concealed human chess-master; the arguments leading to this conclusion, however, were frequently fallacious. Poe assumes, for example, that it is as easy to design a machine which will invariably win as one which wins occasionally, and argues that since the Automaton was not invincible it was therefore operated by a human, a clear *non sequitur*. For a complete account of the history and method of operation of the Automaton, the reader is referred to a series of articles by Harkness and Battell in *Chess Review*, 1947.

A more honest attempt to design a chess-playing machine was made in 1914 by Torrès y Quévedo, who constructed a device which played an end game of king and rook against king (Vigneron, 1914). The machine played the side with king and rook and would force checkmate in a few moves however its human opponent played. Since an explicit set of rules can be given for making satisfactory moves in such an end game, the problem is relatively simple, but the idea was quite advanced for that period.

The thesis we will develop is that modern general purpose computers can be used to play a tolerably good game of chess by the use of a suitable computing routine or " program ". While the approach given here is believed fundamentally sound, it will be evident that much further experimental and theoretical work remains to be done.

2. General Considerations.

A chess " position " may be defined to include the following data :—

(1) A statement of the positions of all pieces on the board.

(2) A statement of which side, White or Black, has the move.

(3) A statement as to whether the kings and rooks have moved. This is important since by moving a rook, for example, the right to castle on that side is forfeited.

(4) A statement of, say, the last move. This will determine whether a possible *en passant* capture is legal, since this privilege is forfeited after one move.

(5) A statement of the number of moves made since the last pawn move or capture. This is important because of the 50 move drawing rule. For simplicity, we will ignore the rule of draw after three repetitions of a position.

In chess there is no chance element apart from the original choice of which player has the first move. This is in contrast with card games, backgammon, etc. Furthermore, in chess each of the two opponents has " perfect information " at each move as to all previous moves (in contrast with Kriegspiel, for example). These two facts imply (von Neumann and Morgenstern, 1944) that any given position of the chess pieces must be either :—

(1) A won position for White. That is, White can force a win, however Black defends.

(2) A draw position. White can force at least a draw, however Black plays, and likewise Black can force at least a draw, however White plays. If both sides play correctly the game will end in a draw.

(3) A won position for Black. Black can force a win, however White plays.

This is, for practical purposes, of the nature of an existence theorem. No practical method is known for determining to which of the three categories a general position belongs. If there were chess would lose most of its interest as a game. One could determine whether the initial position is won, drawn, or lost for White and the outcome of a game between opponents knowing the method would be fully determined at the choice of the first move. Supposing the initial position a draw (as suggested by empirical evidence from master games*) every game would end in a draw.

It is interesting that a slight change in the rules of chess gives a game for which it is provable that White has at least a draw in the initial position. Suppose the rules the same as those of chess except that a player is not forced to move a piece at his turn to play, but may, if he chooses, " pass ". Then we can prove as a theorem that White can at least draw by proper play. For in the initial position either he has a winning move or not. If so, let him make this move. If not, let him pass. Black is now faced with essentially the same position that White had before, because of the mirror symmetry of the initial position †. Since White had no winning move before, Black has none now. Hence, Black at best can draw. Therefore, in either case White can at least draw.

---

* The world championship match between Capablanca and Alekhine ended with the score Alekhine 6, Capablanca 3, drawn 25.

† The fact that the number of moves remaining before a draw is called by the 50-move rule has decreased does not affect this argument.

In some games there is a simple *evaluation function* $f(\mathrm{P})$ which can be applied to a position P and whose value determines to which category (won, lost, etc.) the position P belongs. In the game of Nim (Hardy and Wright, 1938), for example, this can be determined by writing the number of matches in each pile in binary notation. These numbers are arranged in a column (as though to add them). If the number of ones in each column is even, the position is lost for the player about to move, otherwise won.

If such an evaluation function $f(\mathrm{P})$ can be found for a game it is easy to design a machine capable of perfect play. It would never lose or draw a won position and never lose a drawn position and if the opponent ever made a mistake the machine would capitalize on it. This could be done as follows: Suppose

$$f(\mathrm{P})=1 \text{ for a won position,}$$
$$f(\mathrm{P})=0 \text{ for a drawn position,}$$
$$f(\mathrm{P})=-1 \text{ for a lost position.}$$

At the machine's turn to move it calculates $f(\mathrm{P})$ for the various positions obtained from the present position by each possible move that can be made. It chooses that move (or one of the set) giving the maximum value to f. In the case of Nim where such a function $f(\mathrm{P})$ is known, a machine has actually been constructed which plays a perfect game *.

With chess it is possible, *in principle*, to play a perfect game or construct a machine to do so as follows: One considers in a given position all possible moves, then all moves for the opponent, etc., to the end of the game (in each variation). The end must occur, by the rules of the game, after a finite number of moves† (remembering the 50 move drawing rule). Each of these variations ends in win, loss or draw. By working backward from the end one can determine whether there is a forced win, the position is a draw or is lost. It is easy to show, however, that even with the high computing speeds available in electronic calculators this computation is impractical. In typical chess positions there will be of the order of 30 legal moves. The number holds fairly constant until the game is nearly finished as shown in Fig. 1. This graph was constructed from data given by De Groot, who averaged the number of legal moves in a large number of master games (De Groot, 1946, *a*). Thus a move for White and then one for Black gives about 10^3 possibilities. A typical game lasts about 40 moves to resignation of one party. This is conservative for our calculation since the machine should calculate out to checkmate, not resignation. However, even at this figure there will be

---

* Condon, Tawney and Derr, U.S. Patent 2,215,544. The "Nimotron" based on this patent was built and exhibited by Westinghouse at the 1938 New York World's Fair.

† The longest possible chess game is 6350 moves, allowing 50 moves between each pawn move or capture. The longest tournament game on record between masters lasted 168 moves, and the shortest four moves. (Chernev, *Curious Chess Facts*, The Black Knight Press, 1937.)

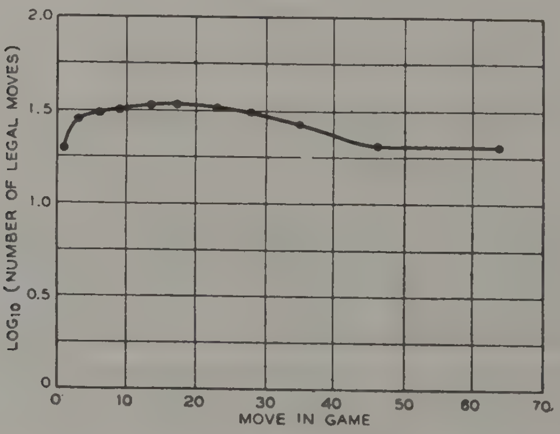
10^{120} variations to be calculated from the initial position. A machine operating at the rate of one variation per micro-microsecond would require over 10^{90} years to calculate its first move !

Another (equally impractical) method is to have a " dictionary " of all possible positions of the chess pieces. For each possible position there is an entry giving the correct move (either calculated by the above process or supplied by a chess master). At the machine's turn to move it merely looks up the position and makes the indicated move. The number of possible positions, of the general order of $64\,! \mid 32\,!\; 8\,!^2\; 2\,!^6$, or roughly 10^{43}, naturally makes such a design unfeasible.

It is clear then that the problem is not that of designing a machine to play perfect chess (which is quite impractical) nor one which merely plays legal chess (which is trivial). We would like it to play a skilful game, perhaps comparable to that of a good human player.

A *strategy* for chess may be described as a process for choosing a move in any given position. If the process always chooses the same move in the same position the strategy is known in the theory of games as a " pure " strategy. If the process involves statistical elements and does not always result in the same choice it is a " mixed " strategy. The following are simple examples of strategies :—

Fig. 1.



(1) Number the possible legal moves in the position P, according to some standard procedure. Choose the first on the list. This is a pure strategy.

(2) Number the legal moves and choose one at random from the list. This is a mixed strategy.

Both of these, of course, are extremely poor strategies, making no attempt to select good moves. Our problem is to develop a tolerably good strategy for selecting the move to be made.

3. APPROXIMATE EVALUATING FUNCTIONS.

Although in chess there is no known simple and exact evaluating function $f(P)$, and probably never will be because of the arbitrary and complicated nature of the rules of the game, it is still possible to perform an approximate evaluation of a position. Any good chess player must, in fact, be able to perform such a position evaluation. Evaluations are based on the general structure of the position, the number and kind of Black and White pieces, pawn formation, mobility, etc. These evaluations are not perfect, but the stronger the player the better his evaluations. Most of the maxims and principles of correct play are really assertions about evaluating positions, for example :—

(1) The relative values of queen, rook, bishop, knight and pawn are about 9, 5, 3, 3, 1, respectively. Thus other things being equal (!) if we add the numbers of pieces for the two sides with these coefficients, the side with the largest total has the better position.

(2) Rooks should be placed on open files. This is part of a more general principle that the side with the greater mobility, other things equal, has the better game.

(3) Backward, isolated and doubled pawns are weak.

(4) An exposed king is a weakness (until the end game).

These and similar principles are only generalizations from empirical evidence of numerous games, and only have a kind of statistical validity Probably any chess principle can be contradicted by particular counter examples. However, from these principles one can construct a crude evaluation function. The following is an example :—

$$f(P)=200(K-K')+9(Q-Q')+5(R-R')+3(B-B'+N-N')+(P-P')$$
$$-\cdot5(D-D'+S-S'+I-I')+\cdot1(M-M')+\ldots$$

in which :—

K, Q, R, B, N, P are the number of White kings, queens, rooks, bishops, knights and pawns on the board.

D, S, I are doubled, backward and isolated White pawns.

M = White mobility (measured, say, as the number of legal moves available to White).

Primed letters are the similar quantities for Black.

The coefficients ·5 and ·1 are merely the writer's rough estimate. Furthermore, there are many other terms that should be included *. The formula is given only for illustrative purposes. Checkmate has been artificially included here by giving the king the large value 200 (anything greater than the maximum of all other terms would do).

It may be noted that this approximate evaluation $f(P)$ has a more or less continuous range of possible values, while with an exact evaluation there are only three possible values. This is as it should be. In practical

---

* See Appendix I.

play a position may be an "easy win" if a player is, for example, a queen ahead, or a very difficult win with only a pawn advantage. In the former case there are many ways to win while in the latter exact play is required, and a single mistake often destroys the advantage. The unlimited intellects assumed in the theory of games, on the other hand, never make a mistake and the smallest winning advantage is as good as mate in one. A game between two such mental giants, Mr. A and Mr. B, would proceed as follows. They sit down at the chessboard, draw for colours, and then survey the pieces for a moment. Then either

(1) Mr. A says, "I resign" or
(2) Mr. B says, "I resign" or
(3) Mr. A says, "I offer a draw," and Mr. B replies, "I accept."

4. STRATEGY BASED ON AN EVALUATION FUNCTION.

A very important point about the simple type of evaluation function given above (and general principles of chess) is that they can only be applied in relatively quiescent positions. For example, in an exchange of queens White plays, say, Q × Q (× = captures) and Black will reply P × Q. It would be absurd to calculate the function $f(\mathrm{P})$ after Q × Q while White is, for a moment, a queen ahead, since Black will immediately recover it. More generally it is meaningless to calculate an evaluation of the general type given above during the course of a combination or a series of exchanges.

More terms could be added to $f(\mathrm{P})$ to account for exchanges in progress, but it appears that combinations, and forced variations in general, are better accounted for by examination of specific variations. This is, in fact, the way chess players calculate. A certain number of variations are investigated move by move until a more or less quiescent position is reached and at this point something of the nature of an evaluation is applied to the resulting position. The player chooses the variation leading to the highest evaluation for him when the opponent is assumed to be playing to reduce this evaluation.

The process can be described mathematically. We omit at first the fact that $f(\mathrm{P})$ should only be applied in quiescent positions. A strategy of play based on $f(\mathrm{P})$ and operating one move deep is the following. Let $\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3, \ldots, \mathrm{M}_s$ be the moves that can be made in position P and let $\mathrm{M}_1\mathrm{P}$, $\mathrm{M}_2\mathrm{P}$, etc. denote symbolically the resulting positions when M_1, M_2, etc. are applied to P. Then one chooses the M_m which maximizes $f(\mathrm{M}_m\mathrm{P})$.

A deeper strategy would consider the opponent's replies. Let M_{i1}, $\mathrm{M}_{i2}, \ldots, \mathrm{M}_{is}$ be the possible answers by Black, if White chooses move M_i. Black should play to *minimize* $f(\mathrm{P})$. Furthermore, his choice occurs *after* White's move. Thus, if White plays M_i Black may be assumed to play the M_{ij} such that

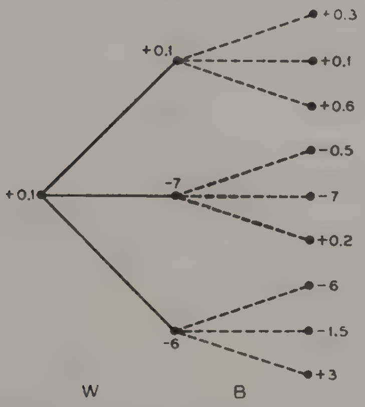
$$f(\mathrm{M}_{ij}\mathrm{M}_i\mathrm{P})$$

is a *minimum*. White should play his first move such that f is a maximum after Black chooses his best reply. Therefore, White should play to maximize on M_i the quantity

$$\min_{M_{ij}} f(M_{ij}M_iP).$$

The mathematical process involved is shown for a simple case in Fig. 2. The point at the left represents the position being considered. It is assumed that there are three possible moves for White, indicated by the three solid lines, and if any of these is made there are three possible moves for Black, indicated by the dashed lines. The possible positions after a White and Black move are then the nine points on the right, and the numbers are the evaluations for these positions. Minimizing on the upper three gives $+\cdot 1$ which is the resulting value if White chooses the upper variation and Black replies with his best move. Similarly, the second and third moves lead to values of -7 and -6. Maximizing on White's move, we obtain $+\cdot 1$ with the upper move as White's best choice.

Fig. 2.



In a similar way a two-move strategy (based on considering all variations out to 2 moves) is given by

$$\max_{M_i}\ \min_{M_{ij}}\ \max_{M_{ijk}}\ \min_{M_{ijkl}} f(M_{ijkl}\ M_{ijk}\ M_{ij}\ M_i\ P) \quad \ldots\ldots (1)$$

The order of maximizing and minimizing this function is important. It derives from the fact that the choices of moves occur in a definite order.

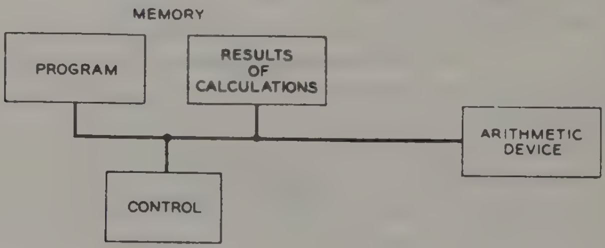
A machine operating on this strategy at the two-move level would first calculate all variations out to two moves (for each side) and the resulting positions. The evaluations $f(P)$ are calculated for each of these positions. Fixing all but the last Black move, this last is varied and the move chosen which minimizes f. This is Black's assumed last move in the variation in question. Another move for White's second move is chosen and the process repeated for Black's second move. This is done for each second White move and the one chosen giving the largest final f (after Black's best assumed reply in each case). In this way White's second move in each variation is determined. Continuing in this way the machine works back to the present position and the best first White move. This move is then played. This process generalizes in the obvious way for any number of moves.

A strategy of this sort, in which all variations are considered out to a definite number of moves and the move then determined from a formula such as (1) will be called a type A strategy. The type A strategy has certain basic weaknesses, which we will discuss later, but is conceptually simple, and we will first show how a computer can be programmed for such a strategy.

5. Programming a General Purpose Computer for a Type A Strategy.

We assume a large-scale digital computer, indicated schematically in Fig. 3, with the following properties :—

Fig. 3.



(1) There is a large internal memory for storing numbers. The memory is divided into a number of boxes each capable of holding, say, a ten-digit number. Each box is assigned a " box number ".

(2) There is an arithmetic organ which can perform the elementary operations of addition, multiplication, etc.

(3) The computer operates under the control of a " program ". The program consists of a sequence of elementary " orders ". A typical order is A 372, 451, 133. This means, extract the contents of box 372 and of box 451, add these numbers, and put the sum in box 133. Another type of order involves a decision, for example, C 291, 118, 345. This tells the machine to compare the contents of box 291 and 118. If the first is larger the machine goes on to the next order in the program. If not, it takes its next order from box 345. This type of order enables the machine to choose from alternative procedures, depending on the results of previous calculations. It is assumed that orders are available for transferring numbers, the arithmetic operations, and decisions.

Our problem is to represent chess as numbers and operations on numbers, and to reduce the strategy decided upon to a sequence of computer orders. We will not carry this out in detail but only outline the programs. As a colleague puts it, the final program for a computer must be written in words of one microsyllable.

The rather Procrustean tactics of forcing chess into an arithmetic computer are dictated by economic considerations. Ideally, we would like to design a special computer for chess containing, in place of the arithmetic organ, a " chess organ " specifically designed to perform the simple chess calculations. Although a large improvement in speed of operation would undoubtedly result, the initial cost of computers seems to prohibit such a possibility. It is planned, however, to experiment with a simple strategy on one of the numerical computers now being constructed.

Fig. 4.

BLACK

70	71	72	73	74	75	76	77
60	61	62	63	64	65	66	67
50	51	52	53	54	55	56	57
40	41	42	43	44	45	46	47
30	31	32	33	34	35	36	37
20	21	22	23	24	25	26	27
10	11	12	13	14	15	16	17
00	01	02	03	04	05	06	07

WHITE

CODE FOR PIECES

	P	N	B	R	Q	K
WHITE	1	2	3	4	5	6
BLACK	-1	-2	-3	-4	-5	-6

0 = EMPTY SQUARE

CODE FOR MOVE

(OLD SQUARE, NEW SQUARE, NEW PIECE (IF PROMOTION))

P - K4 → (14, 34, −)
P - K8(Q) → (64, 74, 5)

A game of chess can be divided into three phases, the opening, the middle game, and the end game. Different principles of play apply in the different phases. In the opening, which generally lasts for about ten moves, development of the pieces to good positions is the main objective. During the middle game tactics and combinations are predominant. This phase lasts until most of the pieces are exchanged, leaving only kings, pawns and perhaps one or two pieces on each side. The end game is mainly concerned with pawn promotion. Exact timing and such possibilities as " Zugzwang ", stalemate, etc. become important.

Due to the difference in strategic aims, different programs should be used for the different phases of a game. We will be chiefly concerned with the middle game and will not consider the end game at all. There seems no reason, however, why an end game strategy cannot be designed and programmed equally well.

A square on a chessboard can be occupied in 13 different ways: either it is empty (0) or occupied by one of the six possible kinds of White pieces (P=1, N=2, B=3, R=4, Q=5, K=6) or one of the six possible Black pieces (P=−1, N=−2, . . ., K=−6). Thus, the state of a square is specified by giving an integer from −6 to +6. The 64 squares can be numbered according to a co-ordinate system as shown in Fig. 4. The position of all pieces is then given by a sequence of 64 numbers each lying between −6 and +6. A total of 256 bits (binary digits) is sufficient memory in this representation. Although not the most efficient encoding, it is a convenient one for calculation. One further number λ will be +1 or −1 according as it is White's or Black's move. A few more should be added for data relating to castling privileges (whether the White or Black kings and rooks have moved), and *en passant* captures (*e.g.*, a statement of the last move). We will neglect these, however. In this notation the starting chess position is given by:—

4, 2, 3, 5, 6, 3, 2, 4; 1, 1, 1, 1, 1, 1, 1, 1;
0, 0, 0, 0, 0, 0, 0, 0; 0, 0, 0, 0, 0, 0, 0, 0;
0, 0, 0, 0, 0, 0, 0, 0; 0, 0, 0, 0, 0, 0, ,0 0;
−1, −1, −1, −1, −1, −1, −1, −1; −4, −2, −3, −5, −6, −3, −2, −4;
+1 (=λ).

A move (apart from castling and pawn promotion) can be specified by giving the original and final squares occupied by the moved piece. Each of these squares is a choice from 64, thus 6 binary digits each is sufficient, a total of 12 for the move. Thus the initial move P—K4 would be represented by 1, 4; 3, 4. To represent pawn promotion a set of three binary digits can be added specifying the piece that the pawn becomes. Castling is described by giving the king move (this being the only way the king can move two squares). Thus, a move is represented by (a, b, c) where a and b are squares and c specifies a piece in case of promotion.

The complete program for a type A strategy consists of nine sub-programs which we designate T_0, T_1, . . ., T_8 and a master program T_9. The basic functions of these programs are as follows:—

T_0—Makes move (a, b, c) in position P to obtain the resulting position.
T_1—Makes a list of the possible moves of a pawn at square (x, y) in position P.
T_2, . . ., T_6—Similarly for other types of pieces: knight, bishop, rook, queen and king.
T_7—Makes list of all possible moves in a given position.
T_8—Calculates the evaluating function f(P) for a given position P.
T_9—Master program; performs maximizing and minimizing calculation to determine proper move.

With a given position P and a move (a, b, c) in the internal memory of the machine it can make the move and obtain the resulting position by the following program T_0.

(1) The square corresponding to number a in the position is located in the position memory.

(2) The number in this square x is extracted and replaced by 0 (empty).

(3) (*a*) If $x=1$, and the first co-ordinate of a is 6 (White pawn being promoted) or if $x=-1$, and the first co-ordinate of a is 1 (Black pawn being promoted), the number c is placed in square b (replacing whatever was there).

(*b*) If $x=6$ and $a–b=2$ (White castles, king side) 0 is placed in squares 04 and 07 and 6 and 4 in squares 06 and 05, respectively. Similarly for the cases $x=6$, $b–a=2$ (White castles, queen side) and $x=-6$, $a–b=\pm 2$ (Black castles, king or queen side).

(*c*) In all other cases, x is placed in square b.

(4) The sign of λ is changed.

For each type of piece there is a program for determining its possible moves. As a typical example the bishop program, T_3, is briefly as follows. Let (x, y) be the co-ordinates of the square occupied by the bishop.

(1) Construct $(x+1, y+1)$ and read the contents u of this square in the position P.

(2) If $u=0$ (empty) list the move (x, y), $(x+1, y+1)$ and start over with $(x+2, y+2)$ instead of $(x+1, y+1)$.

If λu is positive (own piece in the square) continue to 3.

If λu is negative (opponent's piece in the square) list the move and continue to 3.

If the square does not exist continue to 3.

(3) Construct $(x+1, y-1)$ and perform similar calculation.

(4) Similarly with $(x-1, y+1)$.

(5) Similarly with $(x-1, y-1)$.

By this program a list is constructed of the possible moves of a bishop in a given position P. Similar programs would list the moves of any other piece. There is considerable scope for opportunism in simplifying these programs; *e. g.*, the queen program, T_5, can be a combination of the bishop and rook programs, T_3 and T_4.

Using the piece programs $T_1 \ldots T_6$ and a controlling program T_7 the machine can construct a list of *all* possible moves in any given position P. The controlling program T_7 is briefly as follows (omitting details) :—

(1) Start at square 1,1 and extract contents x.

(2) If λx is positive start corresponding piece program T_x and when complete return to (1) adding 1 to square number. If λx is zero or negative, return to 1 adding 1 to square number.

(3) Test each of the listed moves for legality and discard those which are illegal. This is done by making each of the moves in the position P (by program T_0) and examining whether it leaves the king in check.

With the programs $T_0 \ldots T_7$ it is possible for the machine to play legal chess, merely making a randomly chosen legal move at each turn to move. The level of play with such a strategy is unbelievably bad.* The writer played a few games against this random strategy and was able to checkmate generally in four or five moves (by fool's mate, etc.). The following game will illustrate the utter purposelessness of random play :—

	White (Random)	Black
(1)	P–KN3	P–K4
(2)	P–Q3	B–B4
(3)	B–Q2	Q–B3
(4)	N–QB3	Q × P mate

We now return to the strategy based on an evaluation $f(P)$. The program T_8 performs the function of evaluating a position according to the agreed-upon $f(P)$. This can be done by the obvious means of scanning the squares and adding the terms involved. It is not difficult to include terms such as doubled pawns, etc.

The final master program T_9 is needed to select the move according to the maximizing and minimizing process indicated above. On the basis of one move (for each side) T_9 works as follows :—

(1) List the legal moves (by T_7) possible in the present position.

(2) Take the first in the list and make this move by T_0, giving position M_1P.

(3) List the Black moves in M_1P.

(4) Apply the first one giving $M_{11}M_1P$, and evaluate by T_8.

(5) Apply the second Black move M_{12} and evaluate.

(6) Compare, and reject the move with the smaller evaluation.

(7) Continue with the third Black move and compare with the retained value, etc.

(8) When the Black moves are exhausted, one will be retained together with its evaluation. The process is now repeated with the second White move.

(9) The final evaluations from these two computations are compared and the maximum retained.

(10) This is continued with all White moves until the best is selected (*i. e.* the one remaining after all are tried). This is the move to be made.

---

* Although there is a finite probability, of the order of 10^{-75}, that random play would win a game from Botvinnik. Bad as random play is, there are even worse strategies which choose moves which actually *aid* the opponent. For example, White's strategy in the following game : 1. P–KB3, P–K4. 2. P–KN4, Q–R5 mate.

These programs are, of course, highly iterative. For that reason they should not require a great deal of program memory if efficiently worked out.

The internal memory for positions and temporary results of calculations when playing three moves deep can be estimated. Three positions should probably be remembered : the initial position, the next to the last, and the last position (now being evaluated). This requires some 800 bits. Furthermore, there are five lists of moves each requiring about $30 \times 12 = 360$ bits, a total of 1800. Finally, about 200 bits would cover the selections and evaluations up to the present calculation. Thus, some 3000 bits should suffice.

6. IMPROVEMENTS IN THE STRATEGY.

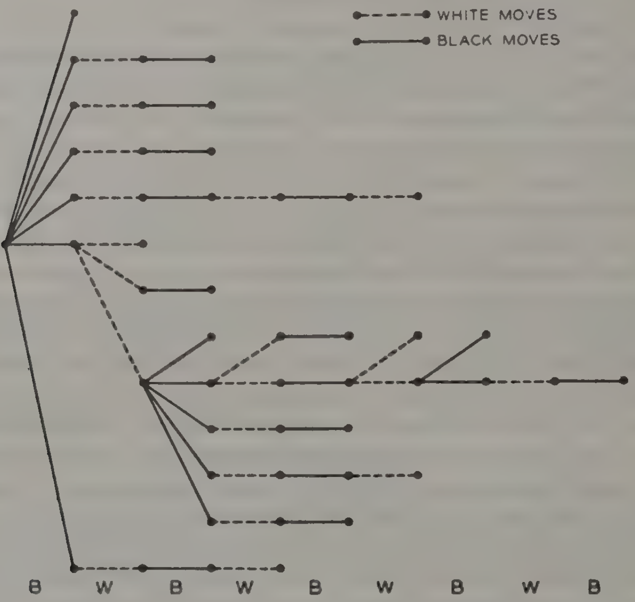
Unfortunately a machine operating according to this type A strategy would be both slow and a weak player. It would be slow since even if each position were evaluated in one microsecond (very optimistic) there are about 10^9 evaluations to be made after three moves (for each side). Thus, more than 16 minutes would be required for a move, or 10 hours for its half of a 40-move game.

It would be weak in playing skill because it is only seeing three moves deep and because we have not included any conditions about quiescent positions for evaluation. The machine is operating in an extremely inefficient fashion—it computes *all* variations to *exactly* three moves and then stops (even though it or the opponent be in check). A good human player examines only a few selected variations and carries these out to a reasonable stopping-point. A world champion can construct (at best) combinations say, 15 or 20 moves deep. Some variations given by Alekhine (" My Best Games of Chess 1924–1937 ") are of this length. Of course, only a few variations are explored to any such depth. In amateur play variations are seldom examined more deeply than six or eight moves, and this only when the moves are of a highly forcing nature (with very limited possible replies). More generally, when there are few threats and forceful moves, most calculations are not deeper than one or two moves, with perhaps half-a-dozen forcing variations explored to three, four or five moves.

On this point a quotation from Reuben Fine (Fine 1942), a leading American master, is interesting : " Very often people have the idea that masters foresee everything or nearly everything ; that when they played P–R3 on the thirteenth move they foresaw that this would be needed to provide a loophole for the king after the complications twenty moves later, or even that when they play 1 P–K4 they do it with the idea of preventing Kt–Q4 on Black's twelfth turn, or they feel that everything is mathematically calculated down to the smirk when the Queen's Rook Pawn queens one move ahead of the opponent's King's Knight's Pawn. All this is, of course, pure fantasy. The best course to follow is to note the major consequences for two moves, but try to work out forced variations as they go."

The amount of selection exercised by chess masters in examining possible variations has been studied experimentally by De Groot (1946, *b*). He showed various typical positions to chess masters and asked them to decide on the best move, describing aloud their analyses of the positions as they thought them through. In this manner the number and depth of the variations examined could be determined. Fig. 5 shows the result of one such experiment. In this case the chess master examined sixteen variations, ranging in depth from 1/2 (one Black move) to 4–1/2 (five Black and four White) moves. The total number of positions considered was 44.

Fig. 5.



From these remarks it appears that to improve the speed and strength of play the machine must :—

(1) Examine forceful variations out as far as possible and evaluate only at reasonable positions, where some quasi-stability has been established.

(2) Select the variations to be explored by some process so that the machine does not waste its time in totally pointless variations.

A strategy with these two improvements will be called a type B strategy. It is not difficult to construct programs incorporating these features. For the first we define a function $g(P)$ of a position which determines whether approximate stability exists (no pieces *en prise*, etc.). A crude definition might be :

$$g(\mathrm{P}) = \begin{cases} 1 \text{ if any piece is attacked by a piece of lower value, or by more} \\ \qquad \text{pieces then defences or if any check exists on a square} \\ \qquad \text{controlled by opponent.} \\ 0 \text{ otherwise.} \end{cases}$$

Using this function, variations could be explored until $g(\mathrm{P})=0$, always, however, going at least two moves and never more than say, 10.

The second improvement would require a function $h(\mathrm{P}, \mathrm{M})$ to decide whether a move M in position P is worth exploring. It is important that this preliminary screening should *not* eliminate moves which merely look bad at first sight, for example, a move which puts a piece *en prise*; frequently such moves are actually very strong since the piece cannot be safely taken.

" Always give check, it may be mate " is tongue-in-cheek advice given to beginners aimed at their predilection for useless checks. " Always investigate a check, it may lead to mate " is sound advice for any player. A check is the most forceful type of move. The opponent's replies are highly limited—he can never answer by counter attack, for example. This means that a variation starting with a check can be more readily calculated than any other. Similarly captures, attacks on major pieces, threats of mate, etc. limit the opponent's replies and should be calculated whether the move looks good at first sight or not. Hence, $h(\mathrm{P}, \mathrm{M})$ should be given large values for all forceful moves (checks, captures and attacking moves), for developing moves, medium values for defensive moves, and low values for other moves. In exploring a variation $h(\mathrm{P}, \mathrm{M})$ would be calculated as the machine computes and would be used to select the variations considered. As it gets further into the variation the requirements on h are set higher so that fewer and fewer subvariations are examined. Thus, it would start considering every first move for itself, only the more forceful replies, etc. By this process its computing efficiency would be greatly improved.

It is believed that an electronic computer incorporating these two improvements in the program would play a fairly strong game, at speeds comparable to human speeds. It may be noted that a machine has several advantages over humans :—

(1) High-speed operation in individual calculations.

(2) Freedom from errors. The only errors will be due to deficiencies of the program while human players are continually guilty of very simple and obvious blunders.

(3) Freedom from laziness. It is all too easy for a human player to make instinctive moves without proper analysis of the position.

(4) Freedom from " nerves ". Human players are prone to blunder due to over-confidence in " won " positions or defeatism and self-recrimination in " lost " positions.

These must be balanced against the flexibility, imagination and inductive and learning capacities of the human mind.

Incidentally, the person who designs the program can calculate the move that the machine will choose in any position, and thus in a sense can play an equally good game. In actual fact, however, the calculation would be impractical because of the time required. On a fair basis of comparison, giving the machine and the designer equal time to decide on a move, the machine might well play a stronger game.

7. Variations in Play and in Style.

As described so far the machine once designed would always make the same move in the same position. If the opponent made the same moves this would always lead to the same game. It is desirable to avoid this, since if the opponent wins one game he could play the same variation and win continuously, due perhaps to some particular position arising in the variation where the machine chooses a very weak move.

One way to prevent this is to have a statistical element in the machine. Whenever there are two or more moves which are of nearly equal value according to the machine's calculations it chooses from them at random. In the same position a second time it may then choose another in the set.

The opening is another place where statistical variation can be introduced. It would seem desirable to have a number of the standard openings stored in a slow-speed memory in the machine. Perhaps a few hundred would be satisfactory. For the first few moves (until either the opponent deviates from the " book " or the end of the stored variation is reached) the machine plays by memory. This is hardly " cheating " since that is the way chess masters play the opening.

It is interesting that the " style " of play of the machine can be changed very easily by altering some of the coefficients and numerical factors involved in the evaluation function and the other programs. By placing high values on positional weaknesses, etc. a positional-type player results. By more intensive examination of forced variations it becomes a combination player. Furthermore, the strength of the play can be easily adjusted by changing the depth of calculation and by omitting or adding terms to the evaluation function.

Finally we may note that a machine of this type will play " brilliantly " up to its limits. It will readily sacrifice a queen or other piece in order to gain more material later or to give checkmate provided the completion of the combination occurs within its computing limits.

The chief weakness is that the machine will not learn by its mistakes. The only way to improve its play is by improving the program. Some thought has been given to designing a program which is self-improving but, although it appears to be possible, the methods thought of so far do not seem to be very practical. One possibility is to have a higher level program which changes the terms and coefficients involved in the evaluation function depending on the results of games the machine has played. Small variations might be introduced in these terms and the values selected to give the greatest percentage of " wins ".

8. ANOTHER TYPE OF STRATEGY.

The strategies described above do not, of course, exhaust the possibilities. In fact, there are undoubtedly others which are far more efficient in the use of available computing time on the machine. Even with the improvements we have discussed the above strategy gives an impression of relying too much on " brute force " calculations rather than on logical analysis of a position. It plays something like a beginner at chess who has been told some of the principles and is possessed of tremendous energy and accuracy for calculation but has no experience with the game. A chess master, on the other hand, has available knowledge of hundreds or perhaps thousands of standard situations, stock combinations, and common manœuvres which occur time and again in the game. There are, for example, the typical sacrifices of a knight at B7 or a bishop at R7, the standard mates such as the " Philidor Legacy ", manœuvres based on pins, forks, discoveries, promotion, etc. In a given position he recognizes some similarity to a familiar situation and this directs his mental calculations along lines with greater probability of success.

There is no reason why a program based on such " type positions " could not be constructed. This would require, however, a rather formidable analysis of the game. Although there are various books analysing combination play and the middle game, they are written for human consumption, not for computing machines. It is possible to give a person one or two specific examples of a general situation and have him understand and apply the general principle involved. With a computer an exact and completely explicit characterization of the situation must be given with all limitations, special cases, etc. taken into account. We are inclined to believe, however, that if this were done a much more efficient program would result.

To program such a strategy we might suppose that any position in the machine is accompanied by a rather elaborate analysis of the tactical structure of the position suitably encoded. This analytical data will state that, for example, the Black knight at B3 is pinned by a bishop, that the White rook at K1 cannot leave the back rank because of a threatened mate on B8, that a White knight at R4 has no move, etc. ; in short, all the facts to which a chess player would ascribe importance in analysing tactical possibilities. These data would be supplied by a program and would be continually changed and kept up-to-date as the game progressed. The analytical data would be used to trigger various other programs depending on the particular nature of the position. A pinned piece should be attacked. If a rook must guard the back rank it cannot guard the pawn in front of it, etc. The machine obtains in this manner suggestions of plausible moves to investigate.

It is not being suggested that we should design the strategy in our own image. Rather it should be matched to the capacities and weaknesses of the computer. The computer is strong in speed and accuracy and weak in analytical ability and recognition. Hence, it should make more use of brutal calculations than humans, but with possible variations increasing by a factor of 10^3 every move, a little selection goes a long way toward improving blind trial and error.

ACKNOWLEDGMENT.

The writer is indebted to E. G. Andrews, L. N. Enequist and H. E. Singleton for a number of suggestions that have been incorporated in the paper.

October 8, 1948.

APPENDIX.

THE EVALUATION FUNCTION FOR CHESS.

The evaluation function $f(P)$ should take into account the " long term " advantages and disadvantages of a position, *i. e.* effects which may be expected to persist over a number of moves longer than individual variations are calculated. Thus the evaluation is mainly concerned with positional or strategic considerations rather than combinatorial or tactical ones. Of course there is no sharp line of division ; many features of a position are on the borderline. It appears, however, that the following might properly be included in $f(P)$:—

(1) Material advantage (difference in total material).
(2) Pawn formation :
  (*a*) Backward, isolated and doubled pawns.
  (*b*) Relative control of centre (pawns at K4, Q4, B4).
  (*c*) Weakness of pawns near king (*e. g.* advanced KNP).
  (*d*) Pawns on opposite colour squares from bishop.
  (*e*) Passed pawns.
(3) Positions of pieces :
  (*a*) Advanced knight (at K5, Q5, B5, K6, Q6, B6), especially if protected by pawn and free from pawn attack.
  (*b*) Rook on open file, or semi-open file.
  (*c*) Rook on seventh rank.
  (*d*) Doubled rooks.
(4) Commitments, attacks and options :
  (*a*) Pieces which are required for guarding functions and, therefore, committed and with limited mobility.
  (*b*) Attacks on pieces which give one player an option of exchanging.
  (*c*) Attacks on squares adjacent to king.
  (*d*) Pins. We mean here immobilizing pins where the pinned piece is of value not greater than the pinning piece ; for example, a knight pinned by a bishop.
(5) Mobility.

These factors will apply in the middle game : during the opening and end game different principles must be used. The relative values to be given each of the above quantities is open to considerable debate, and should be determined by some experimental procedure. There are also numerous other factors which may well be worth inclusion. The more violent tactical weapons, such as discovered checks, forks and pins by a piece of lower value are omitted since they are best accounted for by the examination of specific variations.

REFERENCES.

CHERNEV, 1937, *Curious Chess Facts*, The Black Knight Press.
DE GROOT, A. D., 1946a, *Het Denken van den Schaker* 17–18, Amsterdam ; 1946b. *Ibid.*, Amsterdam, 207.
FINE, R., 1942, *Chess the Easy Way*, 79, David McKay.
HARDY and WRIGHT, 1938, *The Theory of Numbers*, 116, Oxford.
VON NEUMANN and MORGENSTERN, 1944, *Theory of Games*, 125, Princeton.
VIGNERON, H., 1914, *Les Automates*, La Natura.
WIENER, N., 1948, *Cybernetics*, John Wiley.

XXIII. *Propagation of Electromagnetic Disturbances along a Thin Wire in a Horizontally Stratified Medium.*

By B. L. Coleman, B.Sc.,
British Electrical and Allied Industries Research Association *.

[Received October 11, 1949.]

Summary.

It is shown that for a thin wire buried at finite depth in a semi-infinite homogeneous medium and subject to a disturbance at a given frequency, there exists an exponential attenuation of the current in the wire with a propagation constant equal to that of the medium. This result is extended to a medium with a number of layers, and to a wire lying in an interface.

List of Principal Symbols.

All quantities are in M.K.S. rationalized units (Stratton 1941). Timewise variation is as $e^{j\omega t}$, this factor being omitted for brevity.

$\left.\begin{matrix} x, y, z \\ r, \theta, z \end{matrix}\right\}$ =Corresponding cartesian and cylindrical polar coordinate systems.
σ =Conductivity of medium.
μ =Permeability of medium.
ϵ =Permittivity of medium.
ω $=2\pi\times$frequency.
γ =Propagation constant of medium. $[j\omega\mu(\sigma+j\omega\epsilon)]^{\frac{1}{2}}$.
$\Gamma(\omega)$ =Propagation constant of disturbance along the wire, at angular pulsation ω.
a =Radius of wire.
h =Height of wire above interface.
$V(x)$ =Potential at x with respect to that at a point at infinity.
$E(x)$ =Potential gradient at x.
$\vec{E}$ =Electric intensity vector ; components E_x, E_y, E_z.
$\vec{H}$ =Magnetic intensity vector ; components H_x, H_y, H_z.
$\vec{\Pi}$ =Hertzian vector potential ; components Π_x, Π_y, Π_z.
$I(x)$ =Current in wire at x.
$\bar{g}$ =Complex leakage conductance per unit length.
$\bar{Z}$ =Complex internal impedance per unit length.
$J_n(z)$ =Bessel function of first kind of order n (McLachlan 1941).
$K_n(z)$ =Modified Bessel function of second kind of order n^2.
u, f, λ =Arbitrary variables.
α $=(u^2+\gamma^2)^{\frac{1}{2}}$.
p $=2\pi jf$.

The convention is used throughout that the root sign, as in $(u^2+\gamma^2)^{\frac{1}{2}}$, or in $(\lambda^2)^{\frac{1}{2}}$, shall mean that root whose real part is non-negative. Subscripts 0 and 1 refer to the upper and lower media respectively.

---

* Communicated by Dr. S. Whitehead.

§ 1. INTRODUCTION.

THE problem of propagation along a buried wire has received attention in recent years in connection with the lightning protection of overhead transmission lines by buried wires (counterpoise) (Sunde 1940), and in the design of protective methods for buried power and telephone cables (Sunde 1945).

The propagation of surges can be calculated from steady state propagation by the use of operational methods such as the Fourier Integral (Campbell and Foster), so that it is necessary to solve only the latter case. Assume a current at a given frequency to be injected into the wire at a given point ; as it proceeds along the wire it leaks to earth, and an associated system of travelling electromagnetic waves is set up in accordance with Maxwellian theory. If the wire, which is assumed infinite in both directions, is buried at infinite depth in an homogeneous medium, the attenuation of the current follows the law :

$$\mathrm{I}(x)=\pm\tfrac{1}{2}\mathrm{I}(0)\exp\{-\gamma\,|\,x\,|\}, \quad \ldots\ldots\ldots \quad (1)$$

where γ is the propagation constant of the medium. The problem is relatively simple compared with that discussed here, because of the inherent symmetry of the system. The simpler result has been extended to the case of a wire lying in the interface between two media of different characteristics, assuming the conductivity and permittivity of the upper medium (air) are negligible, and the result is

$$\mathrm{I}(x)=\pm\tfrac{1}{2}\mathrm{I}(0)\exp\{-\gamma\,|\,x\,|\,/\sqrt{2}\}, \quad \ldots\ldots \quad (2)$$

where γ is the propagation constant of the lower medium (earth). It has been further suggested (Sunde 1945) that in this case, for moderate depths of burial of the wire, propagation is given approximately by equation (2). However, the method of proof of these results is open to criticism on mathematical grounds.

This paper deals with the general question of propagation along a wire buried at a finite depth in an earth consisting of a number of horizontal stratifications or layers, of which the air can be considered as one, of individual characteristics σ, μ, ϵ, each or all of which may differ between the various layers. It is shown that for thin wires there exists an exponential mode of propagation of the form $\exp\{-\Gamma\,|\,x\,|\}$, where the propagation constant Γ is determined as follows :

(1) If the wire is totally immersed in a layer of constant γ, that is, if it is more than a few diameters away from the nearest interface, then $\Gamma=\gamma$.

(2) If the wire lies at the interface between layers m and $m+1$

$$\Gamma=\left(\frac{\mu_{m+1}\gamma_m^2+\mu_m\gamma_{m+1}^2}{\mu_m+\mu_{m+1}}\right)^{\frac{1}{2}}. \quad \ldots\ldots\ldots \quad (3)$$

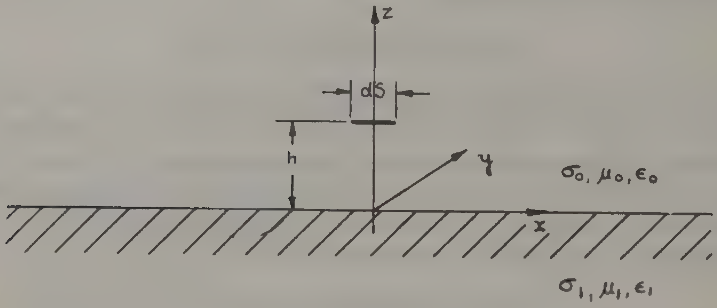
Writing $\gamma_m=0$, $\mu_m=\mu_{m+1}$, equation (2) follows.

The proof as given below is for the case of a two-layer medium, but from its methods, and the nature of the result, the above generalization is immediate (§ 5). It follows the methods used by Sunde (1936, 1940), extending his results and putting them on a rigorous mathematical foundation.

§ 2. THE FIELD OF A DOUBLET.

The doublet is supposed at height h above the interface between two semi-infinite homogeneous media specified by σ_0, μ_0, ϵ_0 and σ_1, μ_1, ϵ_1, as shown in fig. 1. It is of length ds, in the direction of the x axis with its mid-point on the z axis at $(0, 0, h)$, and carrying unit current. If the two media were identical its field would be completely specified by a Hertzian vector potential everywhere in the x direction of magnitude

Fig. 1.



$$\pi'_{0x}=\mathrm{A}\int_0^\infty \frac{\exp\{-\alpha_0|z-h|\}}{\alpha_0}\, u\mathrm{J}_0\,(ru)\,du, \quad . \quad . \quad . \quad . \quad (4)$$

where

$$\mathrm{A}=\frac{j\omega\mu_0\,d\mathrm{S}}{4\pi\gamma_0^2}$$

and (r, θ, z) is the associated cylindrical polar coordinate system. When the media are as shown (4) gives the primary field in the upper region, and the complete field is obtained by superimposing suitable secondary or reflection fields in the upper and lower media. These reflection fields are similarly obtainable from Hertzian vector potentials, and symmetry with respect to the x-z plane restricts such vectors to the x and z directions, so that the complete field can be specified by

$$\pi_{0x}=\pi'_{0x}+\pi''_{0x}, \quad \pi_{0z}$$

$$\pi_{1x} \qquad\quad , \quad \pi_{1z}$$

in the two media. The secondary components are everywhere finite and continuous, since the only permissible singularity is the doublet allowed for in π'_{0x}. The field intensities are given by

$$\left.\begin{aligned} \mathrm{H}_x &= \frac{\gamma^2}{j\omega\mu}\frac{\partial \pi_z}{\partial y}, \\ \mathrm{H}_y &= \frac{\gamma^2}{j\omega\mu}\left(\frac{\partial \pi_x}{\partial z} - \frac{\partial \pi_z}{\partial x}\right), \\ \mathrm{H}_z &= \frac{\gamma^2}{j\omega\mu}\frac{\partial \pi_x}{\partial y}, \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \quad (5)$$

$$\left.\begin{aligned} \mathrm{E}_x &= -\gamma^2\pi_x + \frac{\partial}{\partial x}\left(\frac{\partial \pi_x}{\partial x} + \frac{\partial \pi_z}{\partial z}\right), \\ \mathrm{E}_y &= \frac{\partial}{\partial y}\left(\frac{\partial \pi_x}{\partial x} + \frac{\partial \pi_z}{\partial z}\right), \\ \mathrm{E}_z &= -\gamma^2\pi_z + \frac{\partial}{\partial z}\left(\frac{\partial \pi_x}{\partial x} + \frac{\partial \pi_z}{\partial z}\right), \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \quad (6)$$

The secondary vector components satisfy Poisson's equation

$$\frac{\partial^2\pi}{\partial x^2} + \frac{\partial^2\pi}{\partial y^2} + \frac{\partial^2\pi}{\partial z^2} = \gamma^2\pi, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (7)$$

and the appropriate general solution is (Stratton 1941, §6·9; eq. (55))

$$\pi = \sum_{n=-\infty}^{n=+\infty} \exp\{nj\theta\} \int_0^\infty [g_n(u)\exp\{-\alpha z\} + f_n(u)\exp\{\alpha z\}] u \mathrm{J}_n(ru)\, du. \quad . \quad (8)$$

The boundary conditions at $z=0$ consist of the continuity of the tangential (x, y) components of $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$. From the nature of the expansions (4) and (8), these conditions may be differentiated or integrated with respect to x and y in the plane $z=0$, and since the potentials must vanish at infinity in this plane the "functional constants" of integration must vanish. The conditions may therefore be taken as

$z=0$

$$\frac{\gamma_0^2}{j\omega\mu_0}\pi_{0z} = \frac{\gamma_1^2}{j\omega\mu_1}\pi_{1z}, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (9)$$

$$\frac{\gamma_0^2}{j\omega\mu_0}\frac{\partial \pi_{0x}}{\partial z} = \frac{\gamma_1^2}{j\omega\mu_1}\frac{\partial \pi_{1x}}{\partial z}, \quad . \; . \; . \; . \; . \; . \; . \quad (10)$$

$$\frac{\partial \pi_{0x}}{\partial x} + \frac{\partial \pi_{0z}}{\partial z} = \frac{\partial \pi_{1x}}{\partial x} + \frac{\partial \pi_{1z}}{\partial z}, \quad . \; . \; . \; . \; . \; . \quad (11)$$

$$\gamma_0^2\pi_{0x} = \gamma_1^2\pi_{1x}. \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (12)$$

Taking the appropriate forms of (5), namely

$$\pi''_{0x} = \sum_{n=-\infty}^{n=+\infty} \exp\{nj\theta\} \int_0^\infty f_{0n}(u)\exp\{-\alpha_0 z\} u \mathrm{J}_n(ru)\, du, \quad . \quad (13)$$

etc., together with (4), equations (10) and (12) give

$$\pi''_{0x}=\int_0^\infty \frac{\mathrm{A}\exp\{-\alpha_0 h\}}{\alpha_0}\cdot\frac{\mu_1\alpha_0-\mu_0\alpha_1}{\mu_1\alpha_0+\mu_0\alpha_1}\exp\{-\alpha_0 z\}u\mathrm{J}_0(ru)\,du, \quad . \quad (14)$$

$$\pi_{1x}=\int_0^\infty \frac{2\gamma_0^2}{\gamma_1^2}\mathrm{A}\exp\{-\alpha_0 h\}\cdot\frac{\mu_1}{\mu_1\alpha_0+\mu_0\alpha_1}\exp\{\alpha_1 z\}u\mathrm{J}_0(ru)\,du, \quad . \quad (15)$$

Noting that

$$\frac{\partial}{\partial x}\mathrm{J}_0(ru)=-\cos\theta u\mathrm{J}_1(ru)=-\tfrac{1}{2}(\exp\{j\theta\}+\exp\{-j\theta\})u\mathrm{J}_1(ru), \quad . \quad (16)$$

equations (9) and (11) yield

$$\pi_{0z}=-\frac{\partial}{\partial x}\int_0^\infty \frac{2(\gamma_0^2-\gamma_1^2)\mu_1\mu_0\mathrm{A}\exp\{-\alpha_0 h\}}{(\mu_1\alpha_0+\mu_0\alpha_1)(\gamma_0^2\mu_1\alpha_1+\gamma_1^2\mu_0\alpha_0)}\exp\{-\alpha_0 z\}u\mathrm{J}_0(ru)\,du, \quad . \quad (17)$$

$$\pi_{1z}=-\frac{\partial}{\partial x}\int_0^\infty \frac{2\gamma_0^2(\gamma_0^2-\gamma_1^2)\mu_1^2\mathrm{A}\ \exp\{-\alpha_0 h\}}{\gamma_1^2(\mu_1\alpha_0+\mu_0\alpha_1)(\gamma_0^2\mu_1\alpha_1+\gamma_1^2\mu_0\alpha_0)}\exp\{\alpha_0 z\}u\mathrm{J}_0(ru)\,du. \quad . \quad (18)$$

The electric intensity in the x direction in the plane $z=h$ is given by

$$\mathrm{E}_x=ds[\mathrm{P}(r)+\frac{\partial^2}{\partial x^2}\mathrm{Q}(r)], \quad . \ . \ . \ . \ . \ . \ . \quad (19)$$

where

$$\mathrm{P}(r)=\frac{-j\omega\mu_0}{4\pi}\int_0^\infty\frac{1}{\alpha_0}\left[1+\frac{\mu_1\alpha_0-\mu_0\alpha_1}{\mu_1\alpha_0+\mu_0\alpha_1}\exp\{-2\alpha_0 h\}\right]u\mathrm{J}_0(ru)\,du, \quad . \quad (20)$$

$$\mathrm{Q}(r)=-\frac{1}{\gamma_0^2}\mathrm{P}(r)+\frac{(\gamma_0^2-\gamma_1^2)j\omega\mu_1\mu_0^2}{2\pi\gamma_0^2}\int_0^\infty\frac{\alpha_0\exp\{-2\alpha_0 h\}u\mathrm{J}_0(ru)\,du}{(\mu_1\alpha_0+\mu_0\alpha_1)(\gamma_0^2\mu_1\alpha_1+\gamma_1^2\mu_0\alpha_0)}. \quad . \quad (21)$$

In the general case of a stratified earth with a number of layers, additional terms must be added to the components of the vector potential to represent higher order reflection fields, and such terms must again satisfy Poisson's equation and be free from singularities in the layer or layers containing the dipole. The effect of such terms on the solution is discussed in §5.

§3. PROPAGATION ALONG AN INFINITE WIRE (Sunde 1936).

The wire is supposed embedded in the upper medium at height h from the interface with its axis along the x axis, and the origin of coordinates is taken at the point on it opposite the centre of the applied distrubance (*e. g.* if current is fed directly into the wire the origin is the point of feed). The wire is of radius a and internal impedance $\overline{\mathrm{Z}}$ (complex) per unit length which is supposed constant along its length for a given frequency. To account for imperfect contact it is supposed surrounded by a shell of negligible thickness and of complex leakage conductance $\bar{g}$ per unit length, again supposed constant along its length for a given frequency.

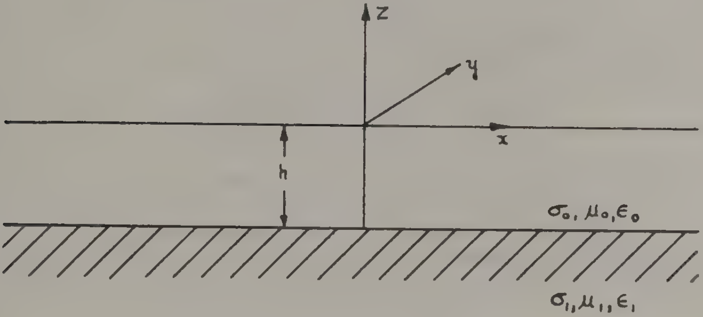
Suppose the applied disturbance gives rise to an electric intensity $\mathrm{E}^0(x)$ along the surface of the wire.

Let $V_\omega(x)$=potential of wire at (x, a, h)
$V_m(x)$=potential of the medium at an adjacent point.
$I(x)$=total current in the wire at x.

Then

$$V_m(x)-V_\omega(x)=\bar{g}^{-1}\frac{dI(x)}{dx}. \qquad (22)$$

Fig. 2.



Let $E_\omega(x)$=potential gradient along the wire at (x, a, h)
$E_m(x)$=potential gradient in the medium at an adjacent point in the x direction.

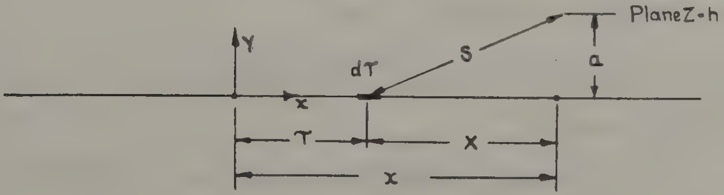
Then

$$E_m(x)-E_\omega(x)=-\bar{g}^{-1}\frac{d^2I(x)}{dx^2}. \qquad (23)$$

Now by definition of $\bar{Z}$ if the wire is sufficiently thin for $E_\omega(x)$ to be constant over its surface, $E_\omega(x)=\bar{Z}I(x)$. Also $E_m(x)$ can be split into two parts, $E^0(x)$ the applied gradient and $E_c(x)$ the gradient due to currents in the surrounding media. Hence

$$\bar{g}^{-1}\frac{d^2I(x)}{dx^2}-\bar{Z}I(x)+E_c(x)=-E^0(x). \qquad (24)$$

Fig. 3.



It remains to express $E_c(x)$ in terms of $I(x)$. With the aid of the results obtained in §2 $E_c(x)$ can be calculated in the plane $z=h$ as follows: referring to fig. 3 consider the effect of a current element at point (τ, o, h)

on the electric field at (x, a, h), and let $X=x-\tau$, $s=(X^2+a^2)^{\frac{1}{2}}$. Then

$$\delta E_c(x)=I(\tau)\,d\tau\left[P(s)+\frac{\partial^2}{\partial X^2}Q(s)\right], \quad . \quad . \quad . \quad . \quad (25)$$

where the P and Q functions are given by (20) and (21).

$$\therefore \quad E_c(x)=\int_{-\infty}^{+\infty}P(s)I(\tau)\,d\tau+\int_{-\infty}^{+\infty}I(\tau)\frac{\partial^2}{\partial X^2}Q(s)\,d\tau$$

$$=T_1+T_2 \text{ (say)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

The current can be expressed by a Fourier Integral as

$$I(x)=\int_{-\infty}^{+\infty}F(\lambda)\exp\{j\lambda x\}\,d\lambda, \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where λ is real on the path of integration. Then

$$*\; T_1=\int_{-\infty}^{+\infty}d\tau P(s)\int_{-\infty}^{+\infty}F(\lambda)\exp\{j\lambda\tau\}\,d\lambda \quad . \quad . \quad . \quad . \quad (28)$$

$$=\int_{-\infty}^{+\infty}d\lambda F(\lambda)\int_{-\infty}^{+\infty}P(s)\exp\{j\lambda\tau\}\,d\tau$$

writing

$$\tau=x+v, \quad b=(v^2+a^2)^{\frac{1}{2}},$$

$$T_1=\int_{-\infty}^{+\infty}d\lambda F(\lambda)\exp\{j\lambda x\}\int_{-\infty}^{+\infty}P(b)\exp\{j\lambda v\}\,dv. \quad . \quad . \quad (29)$$

In a similar manner, noting that $\frac{\partial}{\partial X}=\frac{\partial}{\partial x}$ it follows that

$$T_2=-\int_{-\infty}^{+\infty}d\lambda F(\lambda)\exp\{j\lambda x\}\lambda^2\int_{-\infty}^{+\infty}Q(b)\exp\{j\lambda v\}\,dv, \quad . \quad (30)$$

writing

$$p(\lambda, a)=-\int_{-\infty}^{+\infty}P(b)\exp\{j\lambda v\}\,dv, \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

$$q(\lambda, a)=\int_{-\infty}^{+\infty}Q(b)\exp\{j\lambda v\}\,dv, \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

and expressing $E^0(x)$ as a Fourier integral

$$E^0(x)=\int_{-\infty}^{+\infty}h(\lambda)\exp\{j\lambda x\}\,d\lambda, \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

equation (24) becomes

$$\int_{-\infty}^{+\infty}F(\lambda)\exp\{j\lambda x\}\{\lambda^2[\bar{g}^{-1}+q(\lambda, a)]+[\bar{Z}+p(\lambda, a)]\}\,d\lambda$$

$$=\int_{-\infty}^{+\infty}h(\lambda)\exp\{j\lambda x\}\,d\lambda. \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

---

* For a discussion of the inversion of this and other integrals see Appendix.

Writing

$$\Delta(\lambda)=\lambda^2[\bar{g}^{-1}+q(\lambda, a)]+[\bar{Z}+p(\lambda, a)], \quad . \quad . \quad . \quad . \quad (35)$$

then

$$F(\lambda)=\frac{h(\lambda)}{\Delta(\lambda)}$$

or

$$I(x)=\int_{-\infty}^{+\infty}\frac{h(\lambda)}{\Delta(\lambda)}\exp\{j\lambda x\}\,d\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

From equations (20) and (21), P(b) and Q(b) are integrals of the form

$$N(b)=\int_0^\infty F(u^2)uJ_0(bu)\,du \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

and hence $p(\lambda, a)$, $q(\lambda, a)$ are integrals of the form

$$G=\int_{-\infty}^{+\infty}N[(v^2+a^2)^{\frac{1}{2}}]\exp\{j\lambda v\}\,dv,$$

where as previously remarked, λ is a real variable. In accordance with the notes in the Appendix it follows that

$$\begin{aligned} G&=\int_0^\infty du F(u^2)u\int_{-\infty}^{+\infty}\exp\{j\lambda v\}J_0[u(a^2+v^2)^{\frac{1}{2}}]\,dv \\ &=\int_{|\lambda|}^\infty F(u^2)u\,.\,\frac{2\cos a(u^2-\lambda^2)^{\frac{1}{2}}}{(u^2-\lambda^2)^{\frac{1}{2}}}\,du \\ &=2\int_0^\infty F(u^2+\lambda^2)\cos au\,du \\ &=\int_{-\infty}^{+\infty}\exp\{jau\}F(u^2+\lambda^2)\,du. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)\end{aligned}$$

(These transformations follow from Campbell and Foster, pair 915.4.) Hence

$$\begin{aligned} p(\lambda, a)=\frac{j\omega\mu_0}{4\pi}&\int_{-\infty}^{+\infty}\exp\{jau\}(u^2+\lambda^2+\gamma_0^2)^{-\frac{1}{2}} \\ &\times\left\{1+\frac{\mu_1(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}-\mu_0(u^2+\lambda^2+\gamma_1^2)^{\frac{1}{2}}}{\mu_1(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}+\mu_0(u^2+\lambda^2+\gamma_1^2)^{\frac{1}{2}}}\exp\{-2h(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}\}\right\}du \\ &\quad . \quad . \quad . \quad (39)\end{aligned}$$

with a similar expression for $q(\lambda, a)$, these expressions holding for real values of λ. Consider them now as functions of a complex variable λ, then as $|\lambda|\rightarrow\infty$

$$\begin{aligned} p(\lambda, a)&\rightarrow\frac{j\omega\mu_0}{2\pi}\int_{-\infty}^{+\infty}\exp\{jau\}(u^2+\lambda^2)^{\frac{1}{2}}\left[1+\frac{\mu_1-\mu_0}{\mu_1+\mu_0}\exp\{-2h(u^2+\lambda^2)^{\frac{1}{2}}\right]du \\ &=\frac{j\omega\mu_0}{2\pi}\left[K_0\{(\lambda^2)^{\frac{1}{2}}|a|\}+\frac{\mu_1-\mu_0}{\mu_1+\mu_0}K_0\{(\lambda^2)^{\frac{1}{2}}(a^2+4h^2)^{\frac{1}{2}}\}\right]. \quad . \quad . \quad . \quad (40) \\ &\rightarrow 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)\end{aligned}$$

(See Campbell and Foster, pairs 868 and 558 and McLachlan, list of formulæ No. 111.)

Similarly $q(\lambda, a)\rightarrow 0$ as $\lambda\rightarrow 0$. If now $E^0(x)$ fulfils certain conditions (Whittaker and Watson 1940), equation (36) can be evaluated by the theory of residues. Suppose the poles of the numerator and zeros of the denominator are given by λ_i, λ_k $(i, k=1, 2, 3 \dots)$ where the integrand is now considered as a function of a complex variable λ, and i, k refer to the upper and lower halves of the λ plane respectively, it follows that

$$I(x)=2\pi j\sum_i R_i \qquad (x>0) \quad \dots\dots \quad (42)$$

$$=-2\pi j\sum_k R_k \qquad (x<0) \quad \dots\dots \quad (43)$$

where R_i R_k are the residues of the integrand at λ_i, λ_k respectively. It has been assumed that there are no real poles or roots, but the corresponding modifications are simple, and in any case it is shown below that it is not necessary in general to consider this possibility. (42) and (43) give the formal solutions of the problem and particular cases of interest are discussed in the next section.

§4. THE PRINCIPAL MODE OF PROPAGATION.

(4.1) *Wire not in Interface.*

In this case $h\neq 0$. Consider first the solutions corresponding to the zeros of the function $\Delta(\lambda)$, *i. e.* to the roots of

$$\lambda^2=-\frac{\overline{Z}+p(\lambda, a)}{\overline{g}^{-1}+q(\lambda, a)} \quad \dots\dots\dots \quad (44)$$

$p(\lambda, a)$ is given by (39), which can be written

$$p(\lambda, a)=\frac{j\omega\mu_0}{4\pi}\{2K_0[(\lambda^2+\gamma_0^2)^{\frac{1}{2}}|a|]+T\}, \quad \dots \quad (45)$$

where

$$T=\int_{-\infty}^{+\infty}\frac{\mu_1(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}-\mu_0(u^2+\lambda^2+\gamma_1^2)^{\frac{1}{2}}}{\mu_1(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}+\mu_0(u^2+\lambda^2+\gamma_1^2)^{\frac{1}{2}}}\,\frac{\exp\{-2h(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}\}}{(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}}\,e^{jau}\,du.$$

Suppose a to decrease indefinitely; it is seen by the theory of residues that T remains finite, whilst $K_0[(\lambda^2+\gamma_0^2)^{\frac{1}{2}}|a|]\rightarrow\infty$, and hence as $a\rightarrow 0$, $|p(\lambda, a)|\rightarrow\infty$. Returning to equations (20) and (21) it follows by an exactly similar argument that $q(\lambda, a)\rightarrow p(\lambda, a)/\gamma_0^2$, and supposing $\overline{Z}$ and $\overline{g}^{-1}$ finite it follows that the roots are given by

$$\lambda^2\rightarrow-\gamma_0^2 \quad \dots\dots\dots\dots \quad (46)$$

and the corresponding component of $I(x)$ is given by

$$I(x)\rightarrow I(0)\exp\{-\gamma_0|x|\}, \quad \dots\dots \quad (47)$$

where in this expression $I(0)$ as obtained by letting x tend to zero may be different according as x is +ve or −ve. For example, in the case of current fed into the wire at $x=0$, it is obvious by symmetry that $I(0+)=-I(0-)$. Furthermore, examination of the corresponding residues R_i and R_k shows that $h(\lambda)$ must have a singularity at $\lambda^2=-\gamma_0^2$, and this places an important restriction on the type of excitation which will cause a disturbance in the wire, though it is a restriction which is satisfied in many practical cases (Sunde 1936).

The remaining modes of propagation correspond to the poles of $h(\lambda)$ other than these, and it is seen that as $a \to 0$, $\lambda^2 \neq \gamma_0^2$, propagation is prevented by the infinity assumed by $\Delta(\lambda)$, so that for thin wires propagation will be exponential with a constant given by $\Gamma = \gamma_0$ (see § 1(1)).

(4.2) *Wire in the Interface.*

Similar arguments concerning the possible modes of propagation corresponding to the poles of $h(\lambda)$ at values of λ other than the roots of $\Delta(\lambda)=0$ still apply, and it is only necessary to consider the roots of this equation, viz. (44). In this case $h=0$ and (20) and (21) now give

$$\mathrm{P}(b) = \frac{-j\omega\mu_0\mu_1}{2\pi}\int_0^\infty \frac{1}{\mu_0\alpha_1+\mu_1\alpha_0}\, u\mathrm{J}_0(bu)\, du \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (48)$$

$$\mathrm{Q}(b) = \frac{j\omega\mu_0\mu_1}{2\pi(\gamma_1\mu_0^2-\gamma_0^2\mu_1^2)}\int_0^\infty \left[\frac{\mu_0^2-\mu_1^2}{\mu_1\alpha_0+\mu_0\alpha_1} - \frac{(\gamma_0^2-\gamma_1^2)\mu_1\mu_0}{\gamma_0^2\mu_1\alpha_1+\gamma_1^2\mu_0\alpha_0}\right] u\mathrm{J}_0(bu)\, du. \qquad (49)$$

Proceeding as above it is necessary to evaluate integrals of the form

$$\mathrm{G} = \int_{-\infty}^{+\infty} \exp\{jau\} \frac{1}{\beta_0(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}+\beta_1(u^2+\lambda^2+\gamma_1^2)^{\frac{1}{2}}}\, du, \quad . \; . \qquad (50)$$

where β_0 and β_1 are complex constants. Now

$$\mathrm{G} = \frac{1}{\beta_1^2-\beta_0^2}\int_{-\infty}^{+\infty} \exp\{jau\}\left\{\frac{\beta_1(u^2+\lambda^2+\gamma_1^2)^{\frac{1}{2}}-\beta_0(u^2+\lambda^2+\gamma_0^2)^{\frac{1}{2}}}{u^2+\lambda^2+\psi^2}\right\} du,$$

where

$$\psi^2 = \frac{\beta_1^2\gamma_1^2-\beta_0^2\gamma_0^2}{\beta_1^2-\beta_0^2}.$$

But

$$\int_{-\infty}^{+\infty} \exp\{jau\} \,.\, \frac{(u^2+\lambda^2+\gamma^2)^{\frac{1}{2}}}{u^2+\lambda^2+\psi^2}\, du = 2\pi\int_{-\infty}^{+\infty} \exp\{2\pi jaf\}\, \frac{(\lambda^2+\gamma^2-p^2)^{\frac{1}{2}}}{\lambda^2+\psi^2-p^2},$$

where $p = 2\pi jf$, and it can be shown (Campbell and Foster, pairs 558, 438, 439, 202) that

$$\lim_{a\to 0}\int_{-\infty}^{+\infty} \exp\{2\pi jaf\}\, \frac{(\lambda^2+\gamma^2-p^2)^{\frac{1}{2}}}{\lambda^2+\psi^2-p^2}\, df = \lim_{a\to 0}\frac{1}{\pi}\,\mathrm{K}_0[\lambda^2+\gamma^2)^{\frac{1}{2}}\,|\,a\,|\,] + \ldots$$

$$+\frac{\gamma^2-\psi^2}{\pi(\lambda^2+\psi^2)^{\frac{1}{2}}}\int_0^\infty \exp\{-x(\lambda^2+\psi^2)^{\frac{1}{2}}\}\mathrm{K}_0[(\lambda^2+\gamma^2)^{\frac{1}{2}}x]\, dx, \quad . \; . \qquad (51)$$

and since the square root has a non-negative real part the second term is finite (Admiralty Computing Service). It follows that as $a \to 0$

$$p(\lambda, a) \to \frac{+j\omega\mu_0\mu_1}{2\pi}\cdot\frac{2}{\mu_0^2-\mu_1^2}\{\mu_0\mathrm{K}_0[a(\lambda^2+\gamma_1^2)^{\frac{1}{2}}]-\mu_1\mathrm{K}_0[a(\lambda^2+\gamma_0^2)^{\frac{1}{2}}]\} \quad . \qquad (52)$$

$$q(\lambda, a) \to \frac{+j\omega\mu_0\mu_1}{2\pi}\cdot\frac{2}{\gamma_1^2\mu_0^2-\gamma_0^2\mu_1^2}\{[\mu_0\mathrm{K}_0\{a(\lambda^2+\gamma_1^2)^{\frac{1}{2}}\}-\mu_1\mathrm{K}_0\{a(\lambda^2+\gamma_0^2)^{\frac{1}{2}}\}]$$

$$+\frac{(\gamma_0^2-\gamma_1^2)\mu_1\mu_0}{\gamma_1^4\mu_0^2-\gamma_0^4\mu_1^2}[\gamma_0^2\mu_1\mathrm{K}_0\{a(\lambda^2+\gamma_1^2)^{\frac{1}{2}}\}-\gamma_1^2\mu_0\mathrm{K}_0\{a(\lambda^2+\gamma_0^2)^{\frac{1}{2}}\}]\}. \quad . \; . \qquad (53)$$

Now as $a \to 0$, $K_0[a(\lambda^2+\gamma^1)^{\frac{1}{2}}] \to \log a$, and effecting this replacement it follows from (44) that

$$\lambda^2 \to -\frac{\gamma_0^2\mu_1+\gamma_1^2\mu_0}{\mu_1+\mu_0}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

so that

$$I(x) \to I(0)\exp\left\{-\left(\frac{\gamma_0^2\mu_1+\gamma_1^2\mu_0}{\mu_1+\mu_0}\right)^{\frac{1}{2}}|x|\right\} \quad . \quad . \quad . \quad . \quad (55)$$

(see remarks following equation (47)), which is the required result as set out in § 1 (2).

§ 5. Extension to Multilayer Medium.

The terms in the integrals (20) and (21) are of two kinds, those involving an exponential factor and those not. The former give rise in subsequent integrals to terms which remain finite as $a \to 0$, whilst terms derived from the latter tend to infinity. In order to extend the theory to multilayer media it is necessary only to evaluate the new P and Q functions. The method given in § 2 is readily extended to this case, and it is seen that the terms in the new expressions for P and Q which represent the effect of distant layers all contain exponential factors, so that following through the analysis given in §§ 3 and 4 the necessary generalization as set out in § 1 is established.

§ 6. Conclusions.

Summarizing the preceding results it is found that for a thin wire there exists an exponential mode of propagation the propagation constant of which is that of the medium in which it is buried ; whilst if it lies in an interface its propagation constant is closely related to those of the two abutting media. In either case distant layers have no effect. This mode is termed the principal mode.

For a wire of finite radius there exists a mode of propagation closely approximating to the principal mode. The change in propagation constant from the interface value to the totally buried value may be expected to be quite sharp, the transition taking full effect at depths of burial of a few diameters, the argument being suitably modified when other layers are present, but in any case distant layers have no effect on the mode of propagation.

Bibliography.

Admiralty Computing Service, *Dictionary of Laplace Transforms*, Pt. 2B, Section 2.3.

Campbell, G. A., and Foster, R. M., *Bell Telephone System Monograph* B584, *Fourier Integrals for Practical Applications.*

McLachlan, N. W., 1941, *Bessel Functions for Engineers* (Oxford : University Press).

Stratton, J. A., 1941, *Electromagnetic Theory* (New York : McGraw-Hill & Co.).

Sunde, E. D., 1936, *Electrical Engineering*, 1338 ; 1940, *Electrical Engineering Trans. Sup.*, 987 ; 1945, *Bell System Tech. J.*, 253.

Whittaker, E. T., and Watson, G. N., 1940, *Modern Analysis*, (Cambridge : Press). Section 6.22 *et seq.*

APPENDIX.

ON THE INTERCHANGE IN THE ORDER OF INTEGRATION IN CERTAIN REPEATED INTEGRALS.

(*a*) In equations (28) and (38) repeated infinite integrals were transformed by interchanging the order of integration. This is always permissible if one pair of limits is finite, and the integrand is continuous in both variables within the limits of integration, but where both limits are infinite the argument breaks down, and it is only in very restricted circumstances that general theorems are available.

In the case of equation (38), where the integrand is a known function, a laborious proof of the legitimacy of the interchange can be constructed, and is outlined in (*b*) below. The justification of (28) must rest on less rigid arguments of a physical nature. If the integrands are functions which represent physically possible distributions of potential, and if the final answer must be that of a real physical problem, an answer obtained by interchanging the order of integration is likely to be correct if there is no reasonable alternative solution.

This presupposes some test of the validity of the solution adopted. An example is the equation

$$\begin{aligned} \int_0^\infty \mathrm{J}_0(at)\cos(bt)\,dt &= (a^2-b^2)^{\frac{1}{2}} \text{ for } b<a \\ &= \quad \infty \qquad ,, \;\; b=a \\ &= \quad 0 \qquad ,, \;\; b>a \end{aligned}$$

an equation similar to those in the text. At $b=a$ two possibilities exist. If b approaches a from below, the limit value of infinity is obtained If b approaches a from above the continuation value of the integral by a summation process is zero. The selection of the value at $b=a$ would depend on physical arguments, assuming that the final result must agree with other available evidence. Thus if the quantity could not be infinite then the representation is inexact and a special study of the physical conditions around $b=a$ would have to be made.

In the present instance the results obtained agree satisfactorily with the physical requirements of the problem and with the available experimental evidence, so that the procedure adopted is considered valid.

(*b*) In equation (38) the transformation

$$\begin{aligned} &\int_0^\infty \cos\lambda v\,dv \int_0^\infty \mathrm{F}(u^2)u\mathrm{J}_0[u(v^2+a^2)^{\frac{1}{2}}]\,du \\ &\qquad = \int_0^\infty \mathrm{F}(u^2)u\,du \int_0^\infty \cos\lambda v\mathrm{J}_0[u(v^2+a^2)^{\frac{1}{2}}]\,dv \quad . \quad . \quad . \quad (56) \end{aligned}$$

is used, in which $\mathrm{F}(u^2)$ is any one of certain functions as described in the text. The inner integral on the L.H.S. is the electric intensity of a Hertzian dipole, so that from energy considerations it must be $\mathrm{L}^2\,(0, \infty)$, and so its Fourier transform exists. The R.H.S. has been shown to exist in the text. Writing unity for $F(u^2)u$ in (56) the truth of the modified equation is demonstrable by direct evaluation of the integrals concerned, so that is sufficient to prove (56) for $F(u^2)u-C$ where C is any constant. This must be done for the real and imaginary parts of $F(u^2)$, which can be any one of the following functions:

$$(u^2+\gamma_0^2)^{-\frac{1}{2}}$$

$$\frac{\mu_1(u^2+\gamma_0^2)^{\frac{1}{2}}-\mu_0(u^2+\gamma_1^2)^{\frac{1}{2}}}{\mu_1(u^2+\gamma_0^2)^{\frac{1}{2}}+\mu_0(u^2+\gamma_1^2)^{\frac{1}{2}}}\cdot\frac{\exp\{-2h(u^2+\gamma_0^2)^{\frac{1}{2}}\}}{(u^2+\gamma_0^2)^{\frac{1}{2}}} \qquad (h\geqslant 0)$$

$$\frac{(u^2+\gamma_0^2)^{\frac{1}{2}}\exp\{-2h(u^2+\gamma_0^2)^{\frac{1}{2}}\}}{[\mu_1(u^2+\gamma_0^2)^{\frac{1}{2}}+\mu_0(u^2+\gamma_1^2)^{\frac{1}{2}}][\gamma_0^2\mu_1(u^2+\gamma_1^2)^{\frac{1}{2}}+\gamma_1^2\mu_0(u^2+\gamma_0^2)^{\frac{1}{2}}]}\cdot (h\geqslant o)$$

Introducing the function

$$H(u)=\frac{Re}{Im}\{u[F(u^2)u-C]\}$$

where in each case C is $\lim_{u\to\infty} F(u^2)u$, it can be shown that u_0 exists such that for $u>u_0$ $H(u)$ is a monotonic function increasing or decreasing to zero, and since the sign of $H(u)$ is immaterial the second case is assumed.

Now by using the asymptotic expansion of the Bessel function it can be shown that

$$\int_{x_1}^{x_2} du \int_{y_1}^{y_2} \frac{\cos\lambda v J_0[u(v^2+a^2)^{\frac{1}{2}}]}{u}\,dv$$

is bounded for $x_1>x_0$, $y_1>y_0$ where $\infty\geqslant x_2\geqslant x_1$, $\infty\geqslant y_2\geqslant y_1$.

It follows by applying the second mean value theorem that

$$\left|\int_{x_1}^{x_2}\frac{H(u)}{u}\,du\int_{y_1}^{y_2}\cos\lambda v J_0[u(v^2+a^2)^{\frac{1}{2}}]\,dv\right|<\epsilon$$

for all $x_2\geqslant x_1\geqslant x(\epsilon)$, $y_2\geqslant y_1\geqslant y(\epsilon)$ and the required result follows without further trouble.

ACKNOWLEDGMENTS.

The author wishes to express his thanks to Dr. S. Whitehead, Director of the British Electrical and Allied Industries Research Association, for permission to publish this paper, to Mr. R. H. Golde for his interest and guidance, and to Mr. S. Michaelson for assistance with the Appendix.

XXIV. *Quantal Aspects of Scientific Information.*

By D. M. MacKay,
Wheatstone Laboratory, King's College, London*.

[Received December 15, 1949.]

Summary.

This paper relates to the borderline linking experimental and theoretical physics with mathematical logic, and covers at several points ground which is common to the theory of communication. Its purpose is, therefore, to introduce in language as far as possible common to workers in all these fields, a quantitative approach to the analysis of scientific concepts and the design of experiments.

Many scientific concepts in different fields have a logically equivalent structure. One can abstract from them a logical form which is quite general, and takes on different particular meanings according to the context. Identification of this structure in precise terms leads to a clarification both of experimental principles and of fundamental relations in different fields of physics.

It is suggested that the most fundamental abstract scientific concept is quantal in its communicable aspects. It is defined as " Information-content " and has two features: the *a priori* or structural, and the *a posteriori* or quantitative. To each there corresponds a *quantum* of information, representing the minimal elementary proposition relating respectively to the structural and quantitative aspects of scientific statements.

Scientific statements are regarded as complexes of these elementary propositions. The information-content of a result can be completely represented by means of a vector in a multidimensional space, the dimensionality of the space and the square of the length of the vector indicating respectively the amounts of structural and quantitative information provided, while the orientation of the vector *specifies* the result. The information-content is shown to be fundamentally limited by the number of *conceptual units of space-time* devoted to the experiment, with obvious practical implications. Expressions are also derived measuring the amount of detail possible in a result under different conditions.

From the standpoint here adopted the various uncertainty-relations of physics appear basically as axioms expressing the quantal nature of communicable information, consequent on the use of logical forms; and the quantity *entropy plus information-content* appears as a fundamental

---

* Communicated by Professor J. T. Randall, F.R.S.

invariant of a physical system. The present paper is, however, intended mainly to be explanatory and introductory, and consideration of more fundamental implications is postponed.

A glossary of less-familiar or new terms introduced is appended.

I. *The Formalism of Information Theory.*

§ 1. INTRODUCTION.

WHEN a physicist ventures into the domain of epistemology he is perhaps especially vulnerable to the charge of arbitrariness in his treatment and choice of definitions. This difficulty is particularly acute in connection with a theory of scientific information, for the brimstone flavour of metaphysics hangs over the subject, and adds zest to criticism. The present paper is an introductory account of an approach which seeks to reduce arbitrariness to a minimum by an appeal to logical first principles.

Experimentation abounds with indications that the everyday concepts of science are not the most fundamental. Each time that a compromise has to be struck, say, between the sensitivity and the response-time of a galvanometer, or the noise-level and band-width of an amplifier, or the resolving power and aperture of a microscope, one has an intuitive feeling that in each case some quantity is remaining constant behind all experimental manipulations—something more fundamental than either of the quantities in question. We say that " Nature cannot be cheated " ; and examples of this principle recur throughout the realm of measurement, and not only in microphysics.

Is there not then a way of expressing scientific facts so that in any context a single universal principle can apply ? Presumably in sufficiently fundamental terms such a principle should become obvious. This thought at least was the stimulus to the present investigation, which leads to the definition of an abstract concept common to all scientific statements, and responsible for their logical significance. In terms of this quantitative concept of *Information*, various semi-intuitive principles can be seen to have a precisely definable basis in a general axiom.

The establishment of a basic formalism for this purpose will occupy the first part of the paper. The second will deal in more detail with matters raised by the first, and will outline some of the practical implications of the theory. It is emphasized, however, that any examples given are intended mainly to help readers with different backgrounds to understand the context, and are not meant to anticipate the more thorough treatment of particular applications which it is hoped to give in later papers. In particular references to communication theory are designed mainly to establish the connection between the author's approach and that which has independently been so elegantly developed by various workers in that field (*e. g.* Gabor 1946, Shannon 1948, Wiener 1948).

§ 2. ACQUISITION OF SCIENTIFIC INFORMATION.

It is necessary, with some apology, to begin with a brief analysis of the nature of the scientific method and scientific propositions, in order to introduce our terms and arrive at the minimal form in which scientific facts can be expressed without loss of significance.

A scientific statement may be defined as a precise description of certain events pictured as populating a tract of a coordinate-space. Its essential purpose is the communication of information derived from an *experiment*—an activity in which events are classified. The acquisition of scientific *information* thus involves two distinct problems, each of which has given rise to a somewhat loose appropriation of the term in a technical sense.

Firstly, one must devise apparatus and/or prepare some system of classification, such that an adequate number of independent categories can be defined when describing the result. For example, if fluctuations varying in frequency between 1 and 100 per second are to be observed, the apparatus must be capable of responding in a time of the order of 1/100 second, *i.e.* of giving 100 " independent " readings per second. Or again, if ten shades of colour are recognized as distinct modes of description (*i.e.* coordinate-values on a " colour-axis ") of the members of a population requiring to be classified, then ten columns must be provided in the observer's notebook.

When a chain of apparatus is involved (including the observer) then the differentiating capacity of the least-discriminating link determines the number of independent categories in the result.

There is a sense in which this number, that is to say the number of independent dimensions or " degrees of freedom ", can be regarded as a measure of the information supplied by the experiment. (Its analogue in the case of communication-theory has been so defined by Gabor (1946)). It is, however, inadequate to represent completely the information derived from an experiment, as we shall see. Giving a more general connotation to a term coined by Gabor, we shall define it in due course as the *logon-content* of a result.

Secondly, the experiment must be performed, by using our apparatus (be it galvanometer, microscope, or eye-plus-notebook) to classify events in the chosen tract of coordinate-space. For instance we may record 100 independent values of a quantity as a function of time—the amount of that quantity which is associated with each of the 100 identifying-points provided by our method in the time-tract considered. We may divide a sample of a population into ten colour-groups, determining the fraction associated with each of the ten identifying points on the colour-axis.

Performance of the experiment thus results in the association of a number with each of the " labels "—categories or degrees of freedom—defined by the structure of the experimental method. The usual scientific statement, however, takes the form of an inference having a certain probability deduced from the experimental result. Thus we arrive at a second common use of the term *information*, to signify the source of *confidence in a given number as representative* of the class identified by its label.

As we shall see, it is possible to give a precise numerical significance to this concept also. It is the essential *a posteriori* complement to the *a priori* logon content, for the complete representation of the information derived from an experiment.

So much for the context in which what follows is relevant. We must now begin at the other end, so to speak, by analysing the nature of the scientific statements made to convey the results of an experiment. In this analysis we shall find a basis, free of arbitrariness, for a quantal representation of scientific information.

§ 3. THE QUANTIZATION OF INFORMATION.

(*a*) *Elementary propositions.*

A scientific statement is a logical form based on *limited* data. As such is can be dissected ultimately into a pattern of elementary " atomic " propositions (Sheffer 1913, Wittgenstein 1922). Each atomic proposition states a fact so simple that it cannot be further decomposed, and hence has only the characteristic " true " or " false " ; its existence is its only attribute. Each therefore owes its individuality only to its position in the logical pattern and may be regarded as a kind of link between two entities in the pattern. Ultimately the latter must make contact with the primitive sense-data in terms of which it acquires meaning, but this aspect of the subject does not concern us here. We shall treat as " elementary " the simplest propositions relating to the concepts of measurement and classification. Any debt so incurred to psychology is better settled out of court.

Ideally, a scientific statement is based entirely on *observable* evidence, and the ideal statement which would describe all the information supplied by a particular experiment is presumably reducible to a pattern of independent elementary propositions relating to observations. We should expect therefore to be able to define a measure of *information-content* corresponding to the number of such propositions substantiated by a given experiment. On this view we are led to define one kind of *unit* of information as *that which decides us to add one elementary proposition* to the " frame " (the pattern of propositions) which is logically sufficient to define the results observed.

(*b*) *Quantization of Information.*

The element of *decision* here appears to be fundamental, for it leads to a quite irreducible *quantization* of the action of scientific information. This is forced on us by our use of logical forms—by our apparent incapability of communicating information unambiguously in any other way. Our statements are rigorous ; to speak of an " imperceptible change " in a statement would be meaningless ; the minimum change possible is the addition of one element to the logical pattern ; and a unit of information is simply that which causes us to make one such addition.

Assuming then that the formation of scientific statements is a quantal process, the next task is to identify our quanta in terms of the processes of experimentation studied in § 2. It is here that care is necessary to avoid confusing the two common uses of the term *information*. *Prior* information is presented by knowledge of the experimental procedure ; *posterior* information arises as the latter is carried out. The one defines the structure of the ultimate statement ; the other defines the amount of evidence which it subsumes. Each is quantal in its communicable aspects, though, of course, the respective " quanta " differ qualitatively. It will best serve our purpose to consider first *posterior* information.

§ 4. POSTERIOR OR METRICAL INFORMATION.

(*a*) *Scale-Units.*

The quantal character of posterior information arises from the way in which we describe a scientific measurement. A description of a result is basically a set of instructions enabling the reader to reproduce for himself a conceptual pattern representing the experience of the observer. The scientific discipline recognizes only one type of elementary experience as necessary and sufficient for this purpose ; the most elementary observational proposition asserts the existence of a *coincidence-relation* between two entities. On the other hand we define a *magnitude* by saying that it occupies a certain interval on a scale. Logically this *occupance-relation* between scale-interval and magnitude is a consequence of the existence of coincidence-relations between the ends of the " unknown " and two definable graduation-entities on the scale.

Now the fineness of graduation of a scale, whether physically marked or estimated, is always limited for reasons which will shortly be examined. The point at present is that for every observation there is a minimum separation between neighbouring graduation-entities, B_{n-1}, B_n, B_{n+1}, say, below which either we cannot define or cannot substantiate with probability greater than one-half, a proposition of the form : " A falls into $B_{n-1}-B_n$ and not into B_n-B_{n+1} ". This smallest interval we shall call the *scale-unit* appropriate to the observation.

(*b*) *Metron-content.*

With a scale marked in such " minimum meaningful intervals ", a magnitude can be specified by the number of intervals which it occupies. This number provides an index of information which we can describe as *metrical*, relating to *measurement*. The unit of metrical information, which we shall term a *metron*, is that which enables one elementary interval to be represented as occupied, in the logical pattern to be communicated. In other words each metron specifies one elementary occupance-relation.

Thus what we carry away from a measurement is basically an integer, the number of conceptually separate occupance-relations which have been specified. This integer we shall refer to as the *metron-content* of the result.

There are certain requirements which must be satisfied if the metron-content is to be equated to the number of atomic propositions justified by a given experimental sequence. It must be so defined that the metron-content of two similar but independent sequences is twice that of either alone. The metron-content of a result must be incapable of augmentation by purely logical manipulation, and all complete representations of a given result (*e. g.* in terms of a frequency-spectrum, or a time-function) should have the same metron-content.

If these conditions are satisfied, a result yielding a given metron-content provides a fixed number of logical elements out of which dependent or equivalent statements can be constructed. We can regard the making of these statements as a rearranging of the elements of the logical pattern. A suitable formalism to represent this process will be considered in § 6.

(*c*) *Proper scales.*

It is obvious that not all physical quantities can be represented on scales having a constant scale-unit. For example the square root of a measured quantity has presumably the same metron-content as the quantity itself, since the latter can be derived from it by a purely logical process; but the metron-content cannot be *proportional* to both quantities. Conceptual scales, which we may call *proper scales*, are not necessarily linear in terms of physical magnitude.

The proper scale of a quantity subject to random fluctuations is a good example of this fact. Fisher (1935) justifies his definition of statistical " information " (which is analogous to our metron-content) as a quantity proportional to the reciprocal of variance, by noting that variance depends inversely on the number of samples involved, and is a measure of the uncertainty with which a given sample can be regarded as representative. If the number of samples is multiplied by n, the information provided (assuming independence) should be n times as great. Since the same process reduces the variance by $1/n$, the information provided should therefore be proportional to its reciprocal.

The metron-content i of a single measurement of a randomly fluctuating quantity x having a probable error $\pm\frac{1}{2}\Delta x$ is therefore not $x/\Delta x$ but $i=x^2/(\Delta x)^2$, and the physical scale of x is divided into significant intervals which are non-uniform*. (The scale of x^2, however, is uniformly divided.) This means that we can give a probability p of just $\frac{1}{2}$ to a statement in the form: x (or x^2) occupies i intervals. The figure of $\frac{1}{2}$ implies, not that we are ignorant of the magnitude of x, but that its terminal is as likely to lie within the ith interval as to fall outside it. It also implies that it would be unprofitable to employ a narrower interval in making our statement, for we should then be bound to make it in a form with $p<\frac{1}{2}$: more likely to be false than true.

---

* On the abstract conceptual scale of course only *order* and *number* are significant, so that all intervals are equivalent. It would be meaningless to say they were *equal*.

(d) *Observation and measurement.*

In more complex cases there is a clearer distinction between *observation* and *measurement*. For instance we observe the position of a pointer, and hence measure a voltage. Now the precision of observation certainly sets an upper limit to that of measurement. Basically we observe the distance between the pointer and some reference mark (which must be a point at which another reading has been taken under standard conditions. It is possible to deduce the metron-content of the figure we obtain for this distance, and we could call it the metron-content of the *observation*, but it is not of fundamental significance. In principle it is the reproducibility of the pattern of events symbolized by the statement made which governs the metron-content of a measurement. The greater the metron-content the more detailed a statement can be made with a 50–50 chance of non-contradiction. Thus fluctuations in the readings may be the limiting factor. Indeed one may say that in a well-designed experiment this is always the case (*cf.* the operation of radio search receivers at the " limit of noise "). An experiment is not giving full information unless the metron-content of the observation exceeds that of the measurement.

(e) *Displacement of origin.*

An important point has not so far been raised. A measured value of a quantity x is inevitably represented as occupying an interval Δx. Accordingly since metron-content is essentially positive, the smallest " observable value " which we can justifiably attribute to a quantity *linearly related* to metron-content is not zero, but $\frac{1}{2}\Delta x$. The first interval which it can occupy on our conceptual scale has a width Δx. A detailed consideration of this point, and of the relation between different information scales, will, however, require a separate paper.

(f) *Summary.*

The suggestion here made may now be summarized as follows : full justice is done to the accuracy (the logical content) of any single measurement when it is described in terms of its *scale-unit* and its *metron-content*. The latter may be represented as the number of intervals, on an abstract conceptual " proper-scale ", occupied by the measured quantity ; but the number of intervals on this scale will not necessarily be proportional to the measured quantity, since the scale-unit may also depend on the latter. Quantities for which such a proportionality does hold, have a special claim to be regarded as fundamental in the particular context.

§ 5. STRUCTURAL INFORMATION.

(*a*) *Logon-content.*

It was Kant who said: "reason has insight only into that which it produces after a plan of its own." The design of an experiment is essentially the specification *a priori* of a pattern, of categories in terms of which alone the result can be described. All the events of the experiment must find a place in one or other of these, though of course not all categories will necessarily find an exemplar in a given experiment.

Since each independent category enables us to introduce a measure of differentiation—*i. e.* of form or structure—into our account of a result, we can regard knowledge thereof as providing us with *prior* or *structural information.* We can therefore define a unit of structural information or (using Gabor's term (*loc. cit.*)) a *logon*, as that which enables us to *formulate* one independent proposition, describing one independent feature of the result. (Whether when we have formulated the proposition we shall find any metrical information to give it logical content, is another matter.) The amount of structural information in a result, the *logon-content*, is thus the number of independent categories or degrees of freedom precisely definable in its description.

(*b*) *Logon-capacity.*

In many cases structure is defined in terms of a reference-coordinate. For example the density-pattern on a photographic plate can be described by a function of one or more space-coordinates; and the structure of a telephony signal can be specified by a time function. The *logon-capacity* of an experimental method can in such cases be defined as the number of logons which it specifies per unit of coordinate-interval, or coordinate-space if several coordinates are involved. The total number of independent categories or features in the result is then the integral of the logon-capacity over the extent of coordinate-tract occupied.

Thus the logon-capacity of a microscope in a particular region in the focal plane can be defined in *logons/cm.*^{2}, and measures the resolving-power in that region; for suitable test-objects a resolving power in a given direction can also be defined, in logons/cm. The logon-capacity of a galvanometer or a communication-channel is measured in logons per second, and represents the number of (practically) independent readings per second which can be made with the apparatus.

(*c*) *Frequency-response.*

It is useful here to introduce the idea of frequency-response in a somewhat generalized form. With instruments measuring functions of time it is common practice to define performance in terms of their frequency-response to sinusoidal inputs of different time-periodicities. In the same way one can define Fourier-variables associated with other coordinates, in order to assess the performance of other types of instrument.

For example one could use sinusoidal density-patterns of different *space*-periodicities to define the " frequency-response " of a microscope, if one may so extend the use of the term. In each case a precise meaning can be given to the *frequency-bandwidth* of an instrument, as the effective range of input-frequencies (*i. e* of the Fourier-variable) to which it is sensitive. (For example in a microscope this is a space-frequency range normally extending from zero to a value directly proportional to the aperture used. It thus measures the fineness of detail perceptible, in terms of a Fourier-type of analysis of the transparency-function which represents the object. One need hardly say that " frequency " here is quite distinct from the frequency of the light used.) It should be noted that with more than one coordinate the Fourier-variables must be represented in a Fourier-space in which " bandwidth " is defined by a " volume " not necessarily rectangular.*

(*d*) *Bandwidth and logon-capacity.*

The bandwidth of an instrument in the above sense is directly related to its logon-capacity. The relation arises from the well-known uncertainty-principle which can be written

$$\Delta f \,.\, \Delta q \geqslant \mathrm{K}_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

where Δf represents the effective range of frequencies (conjugate to a coordinate q) to which the apparatus is sensitive, Δq twice the " uncertainty "† in q, and K_s a number which we may take to have the value $\frac{1}{2}$.

Equation (2) will later be justified from our present point of view. Meanwhile we may note that points on the q-axis cannot be defined uniquely at closer intervals than $\mathrm{K}_s/\Delta f$, so that the logon-capacity is $\Delta f/\mathrm{K}_s$ or $2\Delta f$. To attempt to talk of " an interval smaller than Δq " would be to try to construct a logical pattern identical with that of " a frequency higher than Δf " which cannot by definition appear in any result and is therefore observationally meaningless. The logon-content l of an experiment involving a tract of extent q is thus

$$l \leqslant q \,.\, \Delta f/\mathrm{K}_s \leqslant 2q \,.\, \Delta f. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

§ 6. REPRESENTATION OF INFORMATION.

(*a*) *The Information Matrix.*

The relation between the two features of information becomes clear in terms of a simple formalism. The definition of information with which we began was *operational*; " information " acquires meaning in terms of what it *does*. Thus to represent the total information, metrical and structural, provided by an experiment, we require a concise representation enabling us to deduce its effect on all conceivable statements relating to the result—in mathematical jargon, an *information-operator*. Suppose that an experiment has yielded l logons. To represent these as independent, we may think of them as defining a set of orthogonal *rays* in an l-dimensional " information-space ", with corresponding unit-vectors $\alpha_1 \alpha_2 \ldots \alpha_l$. Taking these unit-vectors as a basis, we can now represent the total information yielded by an experiment by an *information-matrix* in the " canonical " form :

---

* The author is indebted to a correspondent for raising this point.

† *i. e.* Δq is the effective range of q.

$$\mathrm{I} = \begin{bmatrix} i_1 & \cdot & \cdot & \cdot \\ \cdot & i_2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & i_l \end{bmatrix}$$

where $i_1 \ldots i_l$ are the metron-contents of the logons represented by $\alpha_1 \ldots \alpha_l$ respectively.

Now relative to the logon-basis α_r, any proposed dependent statement can be represented by a unit vector-function ϕ, say, defined by $\phi = \sum\limits_{r=1}^{l} f_r \,.\, \alpha_r$, where the f_r are direction-cosines of ϕ (or its corresponding ray) and measure the relevance of the corresponding logons to the statement. If now for example we wish to calculate the total metron-content i_φ which the proposed statement would have, we need only form the product

$$i_\varphi = \phi' \,.\, \mathrm{I}\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where ϕ' is the transpose of ϕ.

This procedure is common in the statistical estimation of significance, and has its analogue in the calculation of expectation values in quantum mechanics. In fact all the necessary mathematics at this point is at hand in textbooks on quantum theory (see *e. g.* Tolman 1938).

Without at present going into detail, we may note briefly three further points : (*a*) the *trace* of I, $i = \sum\limits_{r=1}^{l} i_r$, represents the total metron-content of the original result ; (*b*) a logically equivalent re-statement of the result corresponds to a *unitary transformation* of the basis, under which I will in general lose its simple diagonal form. This means that each " component " of a dependent statement is no longer a function of just one of the original logons. But (*c*) the trace of the information matrix is invariant under unitary transformations, so that—as our axiom demands—the total metron-content is unaltered by logical reformulation.

(*b*) *Geometrical Representation.*

As in the analogous formalism of quantum theory, this procedure can be given a simple geometrical representation (see *e. g.* Rojansky 1946). Because of its diagonal form, the operator I can be defined by a vector with components $\sqrt{(i_1)}$, $\sqrt{(i_2)}$, $\ldots$ $\sqrt{(i_l)}$, which we can call the information-vector. This has a length $\sqrt{(i)}$, and defines a point, or rather a volume-element, in the information space, relative to the origin. The information-vector, or the element whose occupance it specifies, can be regarded as the basic invariant behind the various statements we may make about the experiment ; it is the " real result ". The components of the vector along each axis are the square roots of the metron-contents of the corresponding logons, in accordance with our conservation-axiom, § 2 (b). Any dependent statement is now defined by a *direction* in the space, and its *metron-content* is the square of the projection of $\sqrt{(i)}$ in that direction. Unitary transformations correspond to *rotations* of the axes, and the significance of points (a) to (c) above is easily seen.

We have seen that the dimensional multiplicity l is fixed by the choice of experimental method (including the volume of space-time occupied). Performance of the experiment results fundamentally in the collection, and allocation to the various logons, of the metron-flow arising from the impact of data on the apparatus plus observer.

The quantal character of this process has one important effect. Since the number of ways in which i metrons can be distributed among l logons is limited, only a certain number of different results can conceivably be given by a particular experiment. An experiment, indeed, is an attempt to choose between a *finite* number of possibilities. If it were not so experimentation would be impossible.

The merit of an experimental method can therefore be gauged, in part at least, by the speed and ease with which it enables one to identify the appropriate information point.

We shall return to the question of the *detail* in a result in Part II., (§ 9).

II. *Practical Details.*

In developing the formalism of Part I. we incurred a number of debts and shelved several problems. These we must seek now to remove in a more detailed discussion.

§ 7. METRICAL SCALE UNITS.

The first omission concerned the nature of the limitations on the accuracy of statements. In either metrical or structural propositions, a numerical magnitude must be quantized in order to be logically representable. The origin of the quantization in the two cases is quite different however, as we shall now show.

(a) *The metrical information-axiom.*

A single metrical statement about a quantity, y say, is logically sterile unless we identify it by a coordinate-label, q say. The proposition : " I have received i metrons relating to y " represents our actual experience. and has no vagueness ; but vagueness arises directly we conceive y to be a function of q, and try to formulate a proposition about y in terms of q.

For example we may say precisely that a certain measurement of power W over a given period is worth 100 metrons. But when we try to give structure to our result—to say how the power varied during that period—we are quite unable to assign finite metron-content to an estimate of the power at any given instant. The most we can do is to speak in terms of a small interval in which at least one metron was gained. The reason is simply that the information *takes time* to accumulate : there is only a finite metron-density on the time axis. (We may remark that there is an equally strong case for saying that *time* requires *information* to make it real, and that our sense of time is an abstraction from the flux of information we receive. This is outside our present subject, however.)

We have now a single logical pattern serving two modes of description. The result is an uncertainty-relation (really an information-axiom) which arises in general as follows :

Metrons, the " atoms " of the information we have received, are " scale-free " That is to say we can construct a number of dependent propositions involving the same number of metrons, and can regard them as logically equivalent statements of the result, as long as we fix appropriate definitions of the *scale-unit* in each case. Thus to translate i metrons into a statement of the magnitude of y we must have a scale-unit $\varDelta y$ such that

$$y=i\,.\,\varDelta y. \qquad (5)$$

If at the same time we wish to say how y varied as a function of q (*i. e.* to spread our information along a q-axis, such as the time-axis) we must do so by defining a *metron-density*, say ρ_i metrons per unit of q ; so that if the information i was gained over a tract $\varDelta q$,

$$i=\int_q^{q+\Delta q}\rho_i(q)\,dq. \qquad (6)$$

If for $\rho_i(q)$ we substitute its average value in the interval $\varDelta q$, and call it ρ_a, we have from (6) : $\rho_a=i\,.\,(1/\varDelta q)$, which we may write

$$\rho_a=i\,.\,\varDelta\rho. \qquad (7)$$

Now the definition (7) of ρ_a is logically equivalent to the definition (5) of y. Both are anchored to the same i metrons. To connect the two statements we define the relation between the corresponding scale-units by means of a conversion-factor K_m such that $\varDelta y=K_m\,.\,\varDelta\rho$. Thus we have the generalized uncertainty-relation or information-axiom

$$\varDelta y\,.\,\varDelta q=K_m, \qquad (8)$$

which shows that the limit to the accuracy of measurement of y is directly proportional to the extent of q devoted to the process.

(*b*) *Natural-units.*

There is however another way of regarding this. We can also write (from (5) and (8))

$$\varDelta q=i\,.\,(K_m/y). \qquad (8a)$$

This is again in the form of (5), and K_m/y can evidently be regarded as a natural unit of q, say ϵ_q, within which just one metron is acquired. In other words, i metrons acquired in an interval Δq enable us conceptually to subdivide it into i intervals of magnitude ϵ_q. Conversely, we can say that *the metron-content of a measurement cannot exceed the ratio of the coordinate tract Δq associated with it, to the appropriate natural-unit of q, ϵ_q.*

ϵ_q can be calculated for any given experiment, and plays the logical role of an atom of space or time. In the most general case in which y is a function both of space and time, the relevant natural unit is the product of all the ϵ_q's, and the total metron-content cannot exceed the number of these space-time units in the total volume of space-time occupied by the experiment. There is a sense in which ϵ_q is a " scale-unit " of q ; but as the tract irreducibly associated with each measurement of y is Δq, Δq is the smallest interval which enters into any proposition involving y, and the symbol Δ is retained with that connotation. In the limit where the metron-content of each measurement is unity, ϵ_q equals Δq.

(*c*) *Practical aspects.*

Space permits only a brief outline of the practical implications of this principle, which provides a ready " criterion of satisfaction " with the accuracy of a given method. It is particularly helpful in considering the statistical matching of one part of an experiment to another. For instance if a weak link in a sequence is known to yield only a certain metron-content i_0, it is possible to estimate the time and/or space which it is worthwhile to devote to each of the remaining links, and to *gain* in overall metron-content per unit of space-time, or *metron-capacity*, by deliberately designing these so as to barter accuracy for speed or compactness. It is wrong to say (*e. g.*) that it " does no harm " to use a galvanometer which is unnecessarily sensitive ; for a less sensitive one could have a more rapid response, and allow several experiments each yielding i_0 to take place in the same time.

(*d*) *Entropy and metron-content.*

The calculation of ϵ_q in particular cases sometimes yields a pleasant sense of having returned to familiar ground by a devious path. For example in measurement of a power W in an environment at a temperature T for which Boltzmann statistics are valid, K_m (equation 8) is the power per metron, sec.$^{-1}$, or the energy per degree of freedom $\frac{1}{2}k$T. ϵ_t, the shortest significant interval of time (corresponding to the acquisition of one metron) is thus kT/2W. The *entropy-change* per metron is evidently

$$\Delta S = \tfrac{1}{2}k \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (9)$$

so that *entropy and metron-content are equivalent quantities, both having quantal aspects*, a change in one being opposite in sign to the change in the other. Without at present going into detail, we may note that in a physics which started from the concept of Information as one of its basic quantities, the sum *Entropy-plus-Information-content* would rank as a fundamental invariant. (See § 10 (*c*) below). From this viewpoint, temperature would appear as a measure of the scale on which energy is represented in the information-pattern.

An analogous relation holds in the case of a space-statistic such as the density of a photographic image.

§ 8. STRUCTURALLY-DEFINED SCALE-UNITS.

(*a*) *The structural Information-axiom.*

In structural propositions the absolute magnitude of y is irrevelant. They are essentially definitions of propositional functions of which y is to be the argument. Accordingly the scale-unit of q can be defined only in terms of coincidence-relations independent of the magnitude of y. In § 5 we used the concept of bandwidth (generalized) to define a scale-unit Δq which was a property of the apparatus used, ascertainable beforehand by an independent experiment, and hence counting as prior information on subsequent occasions. The general relation (2) which we used then, can be justified directly in terms of our initial axiom that only coincidence-relations or compounds thereof are valid as logical elements in scientific statements.

Let us consider then the case of a simple harmonic function of q, with periodicity f. It will be associated with definitive points on the q-axis, independent of amplitude, whenever the function crosses the axis, *i. e.* at intervals of half a period. (It is also arguable that ingenious experimentation should make it possible to define the quarter-period or the radian-period independently of amplitude.)

In any case, the scale of q is conceptually provided with a set of points at intervals of K_s/f, where K_s is of the order 1.

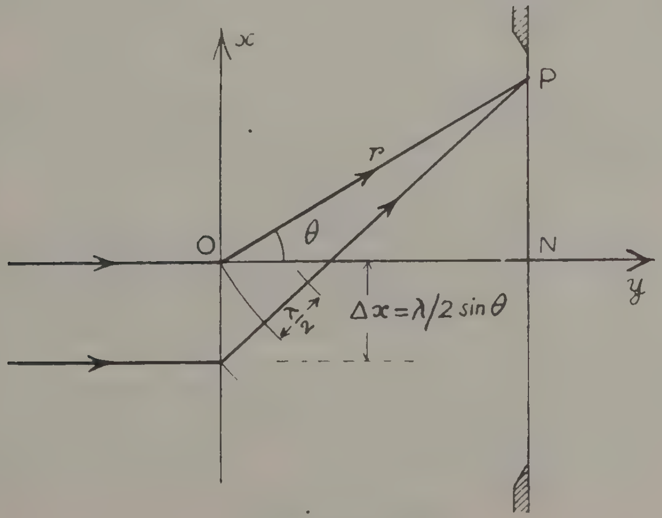
These points, however, are all logically identical, representing collectively a single fact—the accurate value of f. What we want is a set of uniquely identifiable points, to serve as labels. To single out desired points we must provide a comparison-pattern to act as a " pointer ". For instance a frequency $f-\Delta f$ will produce a pattern coinciding with the first at intervals of $K_s/\Delta f$. If all values of Δf from zero can be observed, a continuous range of intervals from infinity down to $K_s/\Delta f$ can be observationally defined. The *structural scale-unit* of q, Δq, is thus $K_s/\Delta f$. Conversely, if Δq is an arbitrary interval, the number l of logons relating to it which can be formulated can be written as $l \leqslant \Delta q \,.\, \Delta f/K_s$, since l is integral. Thus for a single logon $\Delta q \,.\, \Delta f \geqslant K_s$ (equation (2). (It should be remembered that in practice the terms " effective length " and " bandwidth " often have a slightly different numerical connotation from that here employed.)

It is of interest that the metron-density function specified by a single logon, for which equation (2) becomes an equality, is a Gaussian probability-function, since our data are exactly the assumptions made by Gauss in his derivation. (See Appendix.) Analysis of a function into logons is effectively description in terms of superimposed Gaussian functions (Gabor, *loc. cit.*).

(*b*) *Resolving Power.*

To illustrate these ideas we may examine the problem of optical resolving power. As our purpose is only expository we shall consider the somewhat unrealistic case of a narrow rectangular aperture and an ideal object which can be represented by a transparency-function of only one coordinate, x say, parallel to the long edge of the aperture. The situation can then be represented by the two-dimensional diagram of fig. 1. P is a point on the aperture, with coordinates (r, θ) relative to a point O on the object. Light is incident in the direction Oy. The problem has two parts. Firstly, we shall establish a relation between the position of P and the information gained through it about the object; and secondly, we can calculate the logon-capacity for a given type of aperture.

Fig. 1.



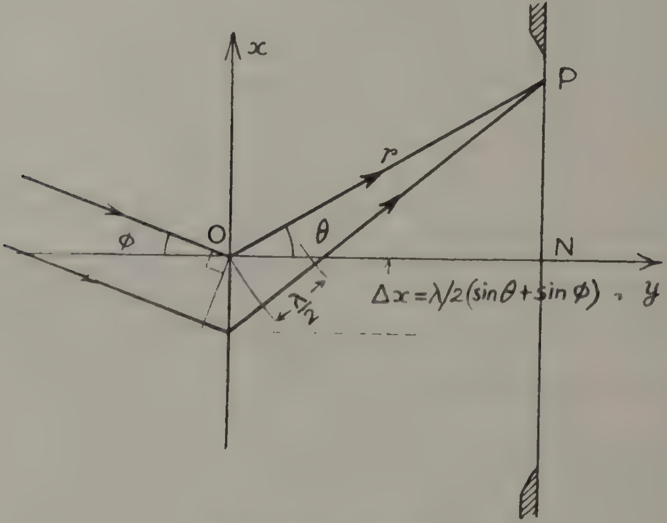
Normal incidence.

Suppose that light of wavelength λ is received through P. This makes it possible to distinguish logically points along the possible light paths to P at intervals of $\lambda/2$, so that with centre P, we can give logical significance to a series of arcs of radii $n\lambda/2$, where n has integral values. These arcs will intersect the x-axis and establish coincidence-relations with the incident wavefront at intervals of $\lambda/2 \sin\theta$, in the vicinity of O.

There is thus a correspondence between the point P and a space-function having a periodicity $f_x = (\sin\theta)/\lambda$ in the object. Indeed if the aperture is closed everywhere except at P and the pole, N, the pattern seen has just this periodicity.

This analysis has assumed that a plane wavefront was incident normally on the x-axis. If instead the light is incident in the xy plane at an angle ϕ to OX (fig. 2) the possible light-paths from P must be traced through the object to their origin on the incident wavefront; and the logically describable paths will be those whose *total* lengths are integral multiples of $\lambda/2$. The corresponding space-periodicity defined on the x-axis will then be $f_x = \left(\frac{1}{\lambda}\right)(\sin\theta + \sin\phi)$.

Fig. 2.



Incidence at angle ϕ.

To calculate the logon-capacity we need only observe that the "bandwidth" is directly proportional to the range of $\sin\theta$. Thus if θ can have all values between $\pm\theta_m$, $\Delta f = \left(\frac{2}{\lambda}\right)(\sin\theta_m)$, and the logon-capacity is $(4/\lambda)(\sin\theta_m)$ *logons per cm.*

This logon-capacity will not, however, be realized in the case of normal incidence, for then f_x is only $(1/\lambda)\sin\theta_m$, and only $(2/\lambda)\sin\theta_m$ points are specified on the x-axis. Only by increasing ϕ to θ_m can the full resolving-power be achieved. Increasing ϕ beyond this value, to give so-called "dark-ground illumination", will increase f_x but not Δf. The number of logons per cm. will not be increased, though owing to the removal of background light the visibility of detail is generally improved.

These results are well-known, but they have here been deduced by way of illustration from an axiom which can perhaps be generalized as follows : *the limit to scientific observation is the limit of our logical vocabulary.* If a phenomenon can be *defined* (in terms of the atomic propositions of the scientific method) it can in principle be *observed.*

§ 9. ESTIMATION OF IMPORTANCE.

(*a*) *Variety.*

We can now calculate the number of logically definable information-points in the l-space of § 6. The number, which we may call M*, measures the total range or variety of the possible results, out of which the experiment has chosen one. It is therefore a measure of the *importance* of a result. For a given upper limit to i, M will depend on the way in which quantization occurs. Assuming that the apparatus is such that all integral values of each i_r can be distinguished, M is easily shown to be

$$\mathrm{M}=(i+l)\,!/i\,!\,l\,! \qquad . \; . \; . \; . \; . \; . \; . \; . \qquad (10)$$

if the total metron-content can have all integral values from 0 to i. If, however, i must always have its maximum value, M is found to be

$$(i+l-1)\,!/i\,!\,(l-1)\,! \qquad . \; . \; . \; . \; . \; . \; . \; . \qquad (11)$$

(This is reasonable, since the possibility of varying i merely confers one further degree of freedom on the system as a whole.)

We can also distinguish between cases in which (*a*) the apparatus sets no limit to the metron-density ρ_i, but only to the total i, and (*b*) the apparatus sets an upper limit to ρ_i. In case (*a*), any i_r can have the full value i, and formula (10) applies. The maximum of M is attained when $l=i$, so that the optimum metron-density in a statement is one metron per logon, as we have seen before. If i is large, Stirling's formula for $i\,!$ then gives

$$\mathrm{M}\simeq 2^{2i}/\sqrt{(\pi i)}. \qquad . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (12)$$

In case (*b*), each i_r can equal but not exceed $\rho_i \,.\, \Delta q$. Interest now focuses on the value of M per unit of q, M_q, which is given by the number of unit " volumes " in an l-dimensional " cube " measuring $(1+\rho_i \,.\, \Delta q)$ in each direction. (The addition of 1 allows for the value $i_r=0$). Since $l=1/\Delta q$ per unit of q, this makes

$$\mathrm{M}_q=(1+\rho_i/l)^l. \qquad . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (13)$$

In the useful limit $l=\rho_i$, so that

$$\mathrm{M}_q=2^l. \qquad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (14)$$

---

* The notation here has been chosen so as to coincide as far as possible with that of Shannon (1949).

(*b*) *Detail.*

A convenient measure of the *detail* in a result is the equivalent number of " yes-no relations ",

$$N=\log_2 M. \qquad \qquad (16)^*$$

For equation (12), in case (*a*), N has the value

$$N=2i-\tfrac{1}{2}\log_2(\pi i) \qquad \qquad (12a)$$

in this case approximately twice the logon- or metron-content. The ability to group all metrons under any one logon approximately doubles the number of independent two-valued propositions which can result from the experiment. In case (*b*), equation (14), the density of detail

$$C=\log_2 M_q=l, \text{ or } \Delta f/K_s. \qquad \qquad (14a)$$

Note, however, that even increasing l to infinity only makes $M_q \rightarrow e^{\rho_i}$, and

$$C \rightarrow \rho_i \log_2 e. \qquad \qquad (15)$$

In some cases (*e. g.* where scale readings are proportional to voltage, in the presence of a constant noise voltage), only whole-number *coordinates* of the information point may be distinguished (and not merely whole-number combinations of metron-content). Mq then cannot exceed

$$(\sqrt{(1+\rho_i/l)})^l, \quad (cf.\ (13)) \quad \text{or} \quad M_q \leqslant \left(1+\frac{\rho_i}{l}\right) l/2. \qquad (17)$$

This is the number of unit-spheres in an l-dimensional sphere of radius $\sqrt{(1+\rho_i/l)}$. When $l=p_i$,

$$M_q=2^{l/2}, \quad \text{and} \quad C=l/2. \qquad \qquad (18)$$

(*c*) *Increase of detail.*

These equations lead to one very practical point. To increase the amount of detail in a result, it is more profitable to increase l than i, unless already $l > i$; but *both methods can be used.*

In other words we have an incentive to apply ingenuity in such matters as the resolving-power of optical instruments, knowing that as long as the metrical information provided is not usefully employed, something practical remains to be done, even if classical theory appears to be discouraging. In the field of communication this aspect is already under vigorous development.

The fundamental point is that as long as ρ_i exceeds the logon-capacity l, there exists a metrical scale-unit smaller than the structural one. One may, therefore, legitimately aspire to formulate in some way or other more than l independent propositions about the structure in question.

---

* This is analogous to Shannon's measure of " information ", and the relations here derived may be compared with those in his paper on *Communication in the presence of Noise* (1949).

§ 10. PRACTICAL ASPECTS.

Thinking in these terms proves helpful and often stimulating in assessing the shortcomings, merits and possibilities of different experimental methods. In particular, subjective preferences in matters of debate can often be usefully supplemented by quantitative considerations of the kind suggested here.

(*a*) *Precision.*

The precision of a single measurement can be enhanced indefinitely by increasing the space-time tract *irreducibly** associated with it. Thus in experiments to determine a constant, efforts should be directed towards " logon-compression "—reducing the frequency-response (*i. e.* the logon-capacity) of the apparatus, with respect to time and space. In short best results are obtained by acting consistently with one's belief that the constant will not alter with time or position, so that one logon will be sufficient. In a sequence of operations, the logon-capacity of each should be adjusted so that the metron-content does not greatly exceed the value which it has in the stage with the narrowest bandwidth. This will enable each subsidiary operation to occupy the minimum space and time, so giving a higher overall metron-capacity, and making possible more repetitions of the experiment in a given space-time tract.

(*b*) *Structural detail.*

In experiments to determine structure on the other hand, the aim should be to reduce the relevant conceptual unit of coordinate to the minimum at which *metron*-content is adequate. One may even say (with reservations) that there is almost always an improvement in resolving power (with a given input of energy) when intelligent steps are taken to sacrifice metrical information. Any device which makes it possible to define a smaller interval can in principle be used to enhance the detail attainable.

(*c*) *Conservation of Information.*

Normally a compromise is required, and the formalism of § 6 is illuminating. Processes of logon-compression and expansion can be represented by rotations of the information-vector. In most processes some attenuation will occur, if they are not reversible in the thermodynamic sense. On the other hand if they are, we deduce from the invariance of entropy plus metron-content that a principle of *conservation of information* applies: *It is impossible for a reversible system to destroy information, however complex may be the process required to reconstitute it in its original form.* At the same time *no artifice of design or manipulation of a result can increase the accuracy of a measurement beyond the limit set by i* (*i. e.* an increase in the metron-content of individual logons can be bought at the expense of logon-capacity, but the limit is set by the total quantity i, which we have seen to depend on the expanse of coordinate-tract devoted to the experiment.

---

* See § 7(*b*).

This in effect is the general principle which lies behind the many " incompatibilities " encountered in experimental design.

(*d*) *Illustrations.*

A typical example of the benefits of logon-compression is the information provided by the analyses of insurance companies. They reduce erratic data obtained in enormous detail, to a small set of propositions of great reliability. An analogous process occurs when a large capacitor is used to smooth the output of a photoelectric cell.

The converse process is well exemplified by the " equalization " of a telephony signal distorted by a narrow-bandwidth channel. Here the effective dimensionality of the information-vector is *increased.* The information-vector is rotated into some of the " dormant " dimensions, which were occupied by larger components in the original signal but had them attenuated in the channel. The consequent all-round reduction in information per logon appears in practice as a corresponding increase in noise.

If, however, the channel is electrically non-dissipative (*i. e.* contains only reactive components), the original signal can theoretically be restored without loss of information.

As a final illustration of the kind of approach here suggested, we may consider the type of aperture in a microscope which would give the maximum of contrast from a particular object. " Contrast " is essentially a matter of metron-content, so that the question may be put : " What transparency-function in the aperture will yield the maximum metron-content ? " We may for illustration make the gross simplification of assuming that contrast is impaired only by random background-light uniformly distributed over the aperture. Again considering the linear case for simplicity, we can describe the object by its " space-spectrum " $S(f_x)$, which will give rise to a corresponding intensity function $S(\xi)$ over the aperture, which extends from ξ_m to $-\xi_m$. If the transparency-function is $t(\xi)$, the total random background power received W_B is proportional to $\int_{-\xi_m}^{\xi} t(\xi)\, d\xi$, while the total coherent power W_C is $\int_{-\xi_m}^{\xi} S(\xi)t(\xi)\, d\xi$. The ratio $\frac{W_C}{W_B}$ will be a maximum when $t(\xi) \equiv S(\xi)$ apart from a constant factor. Thus the aperture should be masked by a transparency representing the spectrum of the expected image. If the requirement is for the maximum resolving-power and metron-content, it is known (see Appendix) that the spectrum corresponding to one logon is Gaussian in outline. Thus the best compromise will be obtained for general purposes by masking the aperture with a Gaussian transparency-function.

§ 11. CONCLUSION.

The chief aim of the present paper has been to present a new way of thinking of experimentation, an effort to isolate the abstract concept which represents the real currency of scientific intercourse, from the various contexts in which it appears. Normal modes of expression of scientific facts involve logical redundancy in vocabulary. Logical analysis in terms of irreducible elements leads to a minimum vocabulary, in which a fact is represented by a conjunction of its abstract information-content with the particular context which gives it concrete meaning.

Towards this minimum vocabulary the present paper can claim to make only a preliminary and naive contribution. It has been seen, however, that scientific information is inherently quantal in its communicable aspects ; and that the various uncertainty-relations of physics, though arising in different ways, are basically expressions of this one fact. *Analogies* in physics emerge as *identities* between basic structures, and the validity of arguments from analogy is easily assessed. It is to be expected that problems analysed in these terms would find an answer with the minimum of logical manipulation.

The concepts here introduced find ready application in the experimental field. Their fruitfulness in other connections remains to be seen. One cannot help feeling, for example, that the dual character of communicable information has a direct bearing on the physiology of thought-processes and the design of electronic reasoning-instruments. Its metrical aspect in particular suggests a possible " analogue " mechanism which is not at present receiving as much attention as the " digital " one, and which it is also hoped to examine in a later paper.

The philosophical reader will recognize that the discussion has not centred on the nature of *reality*, but on the nature of the propositions which the scientific method permits us to make about reality. The implications which this " compulsion by the logical form " have for fundamental theory cannot be examined here ; but if the present approach can help in any way to close familiar loopholes in otherwise attractive epistemological arguments, it will repay further study.

§ 12. ACKNOWLEDGMENTS.

This paper is an expanded version of a lecture given to the Maxwell Society of King's College in January 1948. Since then several papers have appeared, chiefly in American journals, dealing with communication theory in terms of a formalism similar in some respects to that suggested here. This has entailed some duplication in § 9 of the present paper, but it has still seemed worth while to present derivations of the various relations in their original form, since the approach is here somewhat different and more general. In particular it is hoped that controversy over alleged " refutations " of particular definitions (see *e. g.* Tuller 1949) may be allayed by the clarification of the complementary relation between different senses of " Information ".

It is a pleasure to express indebtedness to Dr. D. Gabor, both for contributions which his work has made to the thought of the present paper, and for helpful comments on its presentation. Especial thanks are due to the many colleagues of the author whose detailed criticism has helped to prevent a number of errors and obscurities.

§ 13. GLOSSARY OF TERMS USED.

Bandwidth.—In general terms, the region of Fourier-space to which the output of an instrument is confined. In particular, the effective frequency-range (conjugate to a given coordinate) to which it responds (§ 5(*c*)).

Information.—That which alters the total statement representing "all that is the case" (§ 2).

Information-space.—The space in which independent propositions are represented by orthogonal rays, and their metron-contents by the squares of distances along these rays (§ 6).

Information-vector.—The vector whose components in information-space are the distances just mentioned (§ 6).

Logon.—Unit of structural information, specifying *form* (§ 5(*a*)).

Logon-capacity.—Number of logons per unit of coordinate-space (§ 5(*b*)).

Logon-content.—Number of logons in a statement (§ 5(*a*)).

Metron.—Unit of metrical information, specifying *magnitude* (§ 4(*b*)).

Metron-capacity (*cf.* logon-capacity) (§ 7(*c*)).

Metron-content (*cf.* logon-content).—A measure of the *support* given by the data to a statement (§ 4(*b*)).

Natural unit.—The coordinate-interval over which one metron is acquired (§ 7(*b*).

Proper-scale.—A representational scale on which *magnitude* is represented by a *number* of intervals occupied, though not necessarily linearly related to that number (§ 4(*c*)).

Scale-unit.—The "smallest meaningful interval" on a scale (§ 4(*a*); §§ 7 and 8).

APPENDIX.

The metron-density function specified by a logon.

Gauss based his derivation of the probability function on three assumptions :

(A) The probability of committing an error z in a magnitude X depends only on z.

(B) All values of X are *a priori* equally acceptable ; the probability is not affected by the value of X.

(C) The most probable value of a magnitude X of which we possess n readings, is the arithmetic mean of these readings.

From these it is shown in any text book on probability that the probability of committing an error of between z and $z+dz$ is $h/\sqrt{(\pi)}\exp(-h^2z^2)dz$, where h is a constant denoting the accuracy of the measurement.

Our problem is to derive a function with analogous properties :

(*a*) The metron-density specified by a logon relating to q_0 depends only on the displacement q from q_0.

(*b*) The density-function is not affected by the choice of q_0.

(*c*) The function has a single maximum at the centroid of the metron distribution.

This follows from the facts that : (1) the two possible elementary occupance-relations specified by the logon must each have a probability of $\frac{1}{2}$. There must therefore be as many metrons in support of occupance of the region to the left of q_0, as to that of the right of q_0. (2) Only one *singularity* can be defined by one logon, so that both asymmetry and multiplicity of turning-values is ruled out. (3) The singularity cannot be a minimum, since one-half of the total area under the curve (metron-content) lies between finite limits defined as $\pm\Delta q/2$. We are thus justified in making use of Gauss's result and deducing that the density function is of the form $A\exp(-h^2q^2)$. If we use condition (*c*) (3), that the inter-quartile range is to be Δq, we find from tables that h must be $0{\cdot}941/\Delta q$.

The Fourier transform of this function is also Gaussian, the indices h in the two cases being reciprocal. The frequency spectrum corresponding to a single logon has the same shape as the metron-density function, and could have been derived in the same way.

REFERENCES.

FISHER, R. A., 1935, *The Design of Experiments*, p. 188. London : Oliver and Boyd.

GABOR, D., 1946, *J. Inst. Elec. Engrs.*, **93** (III), 429.

ROJANSKY, V., 1946, *Introductory Quantum Mechanics*, Ch. IX. New York : Prentice-Hall.

SHANNON, C. E., 1948, *Bell System Tech. J.*, **27**, 379–423 ; 623–656 ; 1949, *Proc. I. R. E.*, **37**, 10.

SHEFFER, 1913, *Trans. Amer. Math. Soc.*, **14**, 481–488.

TOLMAN, R. C., 1938, *Principles of Statistical Mechanics.* Oxford University Press.

TULLER, W. G., 1949, *Proc. I. R. E.*, **37**, 468.

WIENER, N., 1948, *Cybernetics.* New York : John Wiley and Sons.

WITTGENSTEIN, 1922, *Tractatus Logico-Philosophicus.* Kegan Paul.

XXV. *Notices of New Books and Periodicals received.*

The Strength of Plastics and Glass. By R. N. HAWARD. [Pp. 245.] (London: Cleaver-Hume Press Ltd., 1949). Price 30*s*. net.

THE scope of this monograph is satisfactorily indicated by its title. The emphasis is mainly on experimental results: indeed in the present state of the theory this could hardly be otherwise. It is primarily as a source of factual information that the book will be valuable, with its fifty tables of data, hundred diagrams, and nearly four hundred references to recent literature. Those not primarily concerned with the technology of the subject will find it a comprehensive summary of the present state of knowledge in a rapidly developing field of investigation. N. T.

Colloid Science, edited by H. R. KRUYT. Volume 2. *Reversible Systems.* (Elsevier Publishing Company, Ltd. 1949). [Pp. 753.] Price £4 10*s*.

DURING the past three decades Professor Kruyt and his school at Utrecht have made contributions of the greatest importance to *Colloid Science*. These contributions have ranged over the whole field of stability of lyophiles, lyophobes, and emulsions, and it is therefore with great pleasure that we welcome this, the first to appear of two large volumes covering many of those topics with which the Dutch school has been concerned. Professor Kruyt is careful to point out in his preface that the book is not comprehensive, but the wide range of topics and the authoritative nature of the treatments by the various expert contributors will make this book essential for all chemistry libraries, and many chemists will want to possess their own copy, in spite of the very high price.

There are seven contributors to this volume. Bungenburg de Jong writes the first chapter, which deals with definitions and general matters like the applicability of the Phase Rule to colloid systems. Modern work on polymers has necessitated a review of our use of many classical terms. De Jong has six more chapters dealing with macromolecular electrolytes, coacervates and complex colloids, thus covering a whole range of water-soluble systems. While the treatment of this section is perhaps rather too extensive in comparison with the others, an accessible account in English has been needed for some time. Also, water-soluble polymers have usually received scant treatment in books on polymer chemistry. Houwink and J. J. and P. H. Hermans are responsible for a very well-documented treatment of non-electrolytic polymers. A strange absentee from the section on osmotic pressure is the well-known apparatus of Fuoss and Mead, which, in fact, is much used by polymer chemists. However, the treatment is otherwise very complete, particularly on the theoretical side. A good account is given of the fundamental theoretical work of Huggins and Flory, and the key position occupied by the experimental work of Gee and Treloar in this field is clearly to be seen.

The book has been very well translated from the Dutch by Dr. L. C. Jackson. It should make a useful companion to the recent volumes under the same title by A. E. Alexander and P. Johnson, to which it is complementary in many respects. The type, illustrations and binding are excellent. We may expect that this timely and thorough account of certain aspects of *Colloid Science*, will greatly stimulate research in the subject. D. D. E.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XXVI. *Theory of Elasticity in General Coordinates.*

By A. E. Green and W. Zerna,
Department of Mathematics, King's College, Newcastle-upon-Tyne [†].

[Received December 22, 1949.]

§ 1. Introduction.

The tensor calculus (Ricci-calculus) has for many years been one of the standard tools in certain branches of mathematics and physics as, for example, in the general theory of relativity. So far, however, only a few writers appear to have used tensor calculus for the presentation of a general theory of elasticity, either for infinitesimal or finite displacements. Elasticity theory is often developed in special rectangular cartesian coordinates and, particularly for finite displacements, the algebraic difficulties are considerable. Also, transformation to any other coordinate system is usually difficult and laborious. The tensor treatment avoids these difficulties and is in many ways easier to follow once the elementary ideas of tensor calculus have been mastered.

In 1937 an important paper was published by Murnaghan (1937), in which he obtained some of the basic relations of the theory of elasticity for finite displacements [‡], using general tensor notations. This seems to be a fundamental paper in many respects but the present writers believe that the theory developed there is not yet in its simplest form. Since 1937 papers have been published [§] in which tensor calculus notations of various types are used for elasticity theory, but the notations and the final forms of some of the results differ considerably, and are not always very simple.

An examination of previous work suggests that this is an appropriate moment in which to attempt a simplification of the presentation of the theory of elasticity in general coordinates. While acknowledging the help derived from previous papers the present writers seek to give a reasonably complete account of elasticity theory in tensor notation, for finite displacements, in which it is believed that the treatment is straightforward, and that the methods used and the relations which are obtained contain many new features.

[†] Communicated by the Authors.

[‡] These relations have recently been expressed in vector and dyadic notations by Milne-Thomson (1948).

[§] Some of these papers are given in the references.

§ 2. DISPLACEMENT VECTOR. DEFORMATION.

Let every point of a continuous three dimensional body, called briefly the body B_0, be at rest, at time $t=0$, relative to a fixed rectangular Cartesian system of axes x_i whose origin is o. The position vector of a typical point P_0 of the body B_0, referred to o as origin, has the form

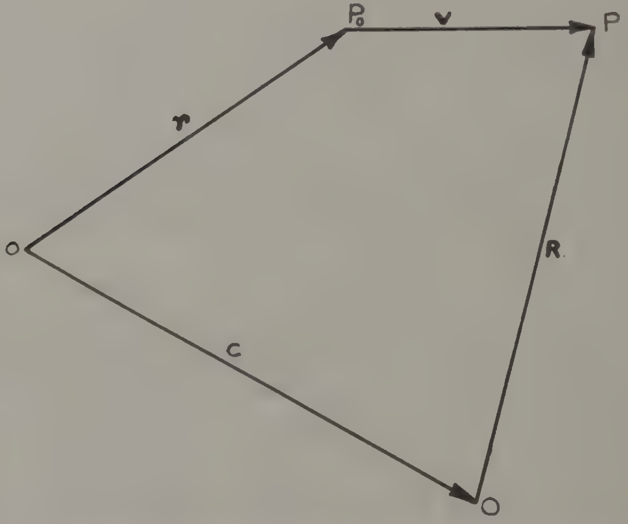
$$\mathbf{r}=\mathbf{r}(x_1, x_2, x_3), \quad \ldots \ldots \quad (2.1)$$

and the line element is given by

$$ds_0^2=dx^i\,dx^i, \quad \ldots \ldots \quad (2.2)$$

where the usual summation convention is used and the latin indices take the values 1, 2, 3.

Fig. 1.



We suppose that the positions of the points of the body B_0 change so that at time t a typical point P_0 has moved to P. The point P may be referred to a new fixed rectangular Cartesian system of axes y_i whose origin is at O, and the position vector of P referred to O is (see fig. 1)

$$\mathbf{R}=\mathbf{R}(y_1, y_2, y_3). \quad \ldots \ldots \quad (2.3)$$

The position vector $\mathbf{v}$ of the point P relative to P_0 represents the displacement of P_0 when the points of B_0 move from their positions at time $t=0$ to their positions at time t. We therefore call

$$\mathbf{v}=\mathbf{R}-\mathbf{r}+\mathbf{c}, \quad \ldots \ldots \quad (2.4)$$

the displacement vector, where the constant vector $\mathbf{c}$ defines the position of the origin O relative to the origin o.

Except for the assumption that the body B_0 is continuous no other assumptions have been made about the material of the body or about the way in which the motion takes place. We now make some assumptions about the way each point P_0 of B_0 moves to its corresponding position P at time t. This will imply that the material of the body has certain physical properties. We shall not, however, discuss the nature of these physical properties. Although we are guided to some extent by physical ideas in making our assumptions, the justification for applying the theory to physical bodies must rest on the comparison of results obtained from theory with those found by experiment.

We wish to study motions of points of the body B_0 which take place continuously and which are such that any curve in B_0 moves so that it always consists of the same particles. In addition, any line element ds_0 of B_0 always remains of the same order of magnitude. We therefore assume that each point P, at time t, is related to its original position P_0, at time $t=0$, by the equations

$$y_i=y_i(x_1, x_2, x_3, t), \qquad \ldots \ldots \qquad (2.5)$$

where y_i are continuous single-valued functions of x_1, x_2, x_3 and t, and where, for each t, these relations are reversible, so that

$$x_i=x_i(y_1, y_2, y_3, t), \qquad \ldots \ldots \qquad (2.6)$$

x_i being single-valued functions of y_1, y_2, y_3 for each value of t.

We also assume that, at a given time t, the differentials dy^i exist and are related to the differentials dx^i by the equations

$$dy^i=\frac{\partial y_i}{\partial x_k}dx^k, \qquad dx^i=\frac{\partial x_i}{\partial y_k}dy^k. \qquad \ldots \ldots \qquad (2.7)$$

These assumptions impose certain restrictions on the partial derivatives of y_i with respect to x_k. Additional restrictions will also be necessary in later parts of the theory but we shall generally suppose, without further discussion, that all conditions which are necessary for the mathematical validity of the theory are satisfied.

If, as a result of the motion of the points of B_0, the relative positions of neighbouring points of the body are altered, so that the length ds_0 of the line element changes to ds, the body B_0 is said to be *strained* or *deformed*. The strained or deformed state of the body, at time t, is denoted by B. The square of the line element in the strained body B may then be written

$$ds^2=dy^i\,dy^i, \qquad \ldots \ldots \qquad (2.8)$$

where the differentials dy^i are to be evaluated for a fixed time t and are related to dx^i by equation (2.7). If $ds=ds_0$ for all points and for every time t the body is said to be rigid.

The original body B_0 may be described by a general curvilinear set of coordinates θ_i so that

$$x_i=x_i(\theta_1, \theta_2, \theta_3). \qquad \ldots \ldots \qquad (2.9)$$

We may imagine that this curvilinear coordinate system moves continuously [†] with the body as we pass from the original state B_0 at time $t=0$ to the state B at time t. It will therefore form a curvilinear system in B so that, from (2.5) and (2.9),

$$y_i=y_i(\theta_1, \theta_2, \theta_3, t). \quad \ldots\ldots\ldots \quad (2.10)$$

Hence, the curvilinear system θ_i may be identified with any convenient coordinate system in either B_0 or B.

We may define in B_0 contravariant components $d\theta^i$ of a vector from the relation (2.9). Thus

$$d\theta^i=\frac{\partial\theta_i}{\partial x_k}dx^k, \qquad dx^k=\frac{\partial x_k}{\partial\theta_i}d\theta^i. \quad \ldots\ldots \quad (2.11)$$

From (2.7) and (2.11) we have

$$d\theta^i=\frac{\partial\theta_i}{\partial y_k}dy^k, \qquad dy^k=\frac{\partial y_k}{\partial\theta_i}d\theta^i, \quad \ldots\ldots \quad (2.12)$$

so that $d\theta^i$ remain the contravariant components of a vector under all transformations, including transformations from coordinate systems in B_0 to those in B.

Using (2.9) and (2.10) the position vectors (2.1) and (2.3) of the points P_0 and P respectively take the forms

$$\left.\begin{array}{r}\mathbf{r}=\mathbf{r}(\theta_1, \theta_2, \theta_3), \\ \mathbf{R}=\mathbf{R}(\theta_1, \theta_2, \theta_3, t),\end{array}\right\}, \quad \ldots\ldots\ldots \quad (2.13)$$

and the displacement vector (2.4) becomes

$$\mathbf{v}=\mathbf{v}(\theta_1, \theta_2, \theta_3, t). \quad \ldots\ldots\ldots \quad (2.14)$$

The line elements (2.2) and (2.8) which correspond to the vectors (2.13) are

$$ds_0^2=g_{ik}\,d\theta^i\,d\theta^k, \quad \ldots\ldots\ldots \quad (2.15)$$

$$ds^2=G_{ik}\,d\theta^i\,d\theta^k, \quad \ldots\ldots\ldots \quad (2.16)$$

where g_{ik} and G_{ik} are the covariant components of the metric tensors of B_0 and B respectively. The quantities G_{ik} are found for a given time t and are functions of t as well as $\theta_1, \theta_2, \theta_3$.

The covariant components g_{ik} of the metric tensor of B_0, together with the corresponding contravariant components g^{ik}, are derived from the Euclidean space defined by the coordinates x_i. These tensor components satisfy the appropriate laws of transformations for all changes of coordinates of the usual type, including changes to coordinate systems in B. Similarly, the covariant and contravariant components G_{ik}, G^{ik} of the metric tensor of B are derived from the Euclidean space defined by the coordinates y_i, and they satisfy the appropriate laws of transformations for all changes of coordinates, including changes to coordinate systems in B_0.

[†] The values of θ_i which define a point P_0 of B_0 remain fixed with P_0 as it moves from its position in B_0 to its position P in B.

Since both B_0 and B are situated in Euclidean spaces it follows that the components of the Riemann–Christoffel tensors for each body vanish, so that

$$ {}_0R_{ijkl}=0, \qquad R_{ijkl}=0, \qquad \ldots\ldots \quad (2.17)$$

the prefix $_0$ referring to the unstrained body B_0. The tensor components ${}_0R_{ijkl}$ are formed from the metric tensor components g_{mn}, g^{mn} and their derivatives with respect to θ_r. Similarly, the tensor components R_{ijkl} are formed from G_{mn} and G^{mn} and their derivatives with respect to θ_r.

§ 3. STRAIN TENSOR.

The state of strain of the body B is characterized[†] by the difference of the squared line elements (2.15) and (2.16). We write

$$ds^2-ds_0^2=2\gamma_{ik}\,d\theta^i\,d\theta^k, \qquad \ldots\ldots \quad (3.1)$$

where

$$\gamma_{ik}=\tfrac{1}{2}(G_{ik}-g_{ik}). \qquad \ldots\ldots\ldots \quad (3.2)$$

The quantities γ_{ik} are the symmetrical covariant components of a tensor which is called the *strain* tensor[‡]. When $\gamma_{ik}=0$ for all points, $ds=ds_0$, and the body is rigid.

Since γ_{ik} are symmetrical it is always possible to find a coordinate system in B such that $\gamma_{ik}=0 (i \neq k)$. Since the coordinate system in B is then orthogonal, $G_{ik}=0$ $(i \neq k)$, and hence, from (3.2), $g_{ik}=0$ $(i \neq k)$, so that the original directions of this coordinate system in B_0 are also orthogonal. The directions of the coordinate curves in B which are determined by this coordinate system are called the principal directions of strain; and the strains in the principal directions, the covariant components of the canonical form of the strain tensor, are termed *principal* strains.

With the help of (3.2) we can express the covariant components of the metric tensor of the body B in terms of the covariant components of the metric tensor of the original body B_0 and the covariant components of the strain tensor. Thus

$$G_{ik}=g_{ik}+2\gamma_{ik} \qquad \ldots\ldots\ldots \quad (3.3)$$

and we obtain the determinant G of the elements G_{ik} in the form

$$\begin{aligned} G&=\tfrac{1}{6}e^{ijk}e^{lmn}G_{il}G_{jm}G_{kn}\\ &=g+\tfrac{1}{3}e^{ijk}e^{lmn}\{g_{il}g_{jm}\gamma_{kn}+g_{jm}g_{kn}\gamma_{il}+g_{il}g_{kn}\gamma_{jm}\\ &\qquad +2(g_{il}\gamma_{jm}\gamma_{kn}+g_{jm}\gamma_{il}\gamma_{kn}+g_{kn}\gamma_{il}\gamma_{jm})+4\gamma_{il}\gamma_{jm}\gamma_{kn}\}, \qquad \ldots \quad (3.4)\end{aligned}$$

[†] The strain may be characterized in other ways as is done, for example, by Richter, H. (1949).

[‡] The strain in the neighbourhood of a point can be interpreted in a geometrical way by a strain " quadric ", or more specially by a " strain ellipsoid " or " reciprocal strain ellipsoid ". See, for example, A. F. H. Love (1927), C. B. Biezeno and K. Grammel (1939).

where g is the determinant of the elements g_{ik} and where e^{ijk} are completely skew-symmetric components of the e-system tensor, which take the values 0, +1, −1.

In order to calculate the strain components in terms of the displacement vector we write (2.4) in the form

$$\mathbf{R}=\mathbf{r}+\mathbf{v}-\mathbf{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.5)$$

and observe that

$$\mathbf{E}_i=\mathbf{e}_i+\mathbf{v}_{,i}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.6)$$

where

$$\mathbf{E}_i=\mathbf{R}_{,i}, \qquad \mathbf{e}_i=\mathbf{r}_{,i}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.7)$$

$\mathbf{e}_i$, $\mathbf{E}_i$ being the covariant base vectors in the original and strained bodies respectively. The comma denotes partial differentiation with respect to θ_i. Hence, since

$$g_{ik}=\mathbf{e}_i \cdot \mathbf{e}_k, \qquad \mathrm{G}_{ik}=\mathbf{E}_i \cdot \mathbf{E}_k, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.8)$$

we have, using (3.6),

$$\mathrm{G}_{ik}=g_{ik}+\mathbf{e}_i \cdot \mathbf{v}_{,k}+\mathbf{e}_k \cdot \mathbf{v}_{,i}+\mathbf{v}_{,i} \cdot \mathbf{v}_{,k},$$

and therefore

$$\gamma_{ik}=\tfrac{1}{2}(\mathbf{e}_i \cdot \mathbf{v}_{,k}+\mathbf{e}_k \cdot \mathbf{v}_{,i}+\mathbf{v}_{,i} \cdot \mathbf{v}_{,k}). \quad . \quad . \quad . \quad . \quad (3.9)$$

The displacement vector may be expressed in terms of the contravariant base vectors $\mathbf{e}^m$ of B_0. Thus

$$\mathbf{v}=v_m\mathbf{e}^m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.10)$$

and

$$\mathbf{v}_{,i}=v_m|_i\mathbf{e}^m, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.11)$$

where the vertical line denotes covariant differentiation with respect to B_0 and means

$$v_m|_i=v_{m,i}-\begin{Bmatrix} r \\ mi \end{Bmatrix}_0 v_r. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.12)$$

The Christoffel symbols $\begin{Bmatrix} r \\ mi \end{Bmatrix}_0$ in the last expression are to be calculated for B_0 from the components g_{mn}, g^{mn} of the metric tensor of B_0 and are given by

$$\left.\begin{aligned} \begin{Bmatrix} r \\ mi \end{Bmatrix}_0 &=g^{rs}[s, mi]_0, \\ [s, mi]_0 &=\tfrac{1}{2}\{g_{si,m}+g_{sm,i}-g_{mi,s}\}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (3.13)$$

Introducing (3.10) into (3.9) the strain components (3.9) become

$$\gamma_{ik}=\tfrac{1}{2}(v_i|_k+v_k|_i+v^r|_iv_r|_k), \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.14)$$

if we use the formulæ

$$\mathbf{e}_i \cdot \mathbf{e}^k=\delta_i^k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.15)$$

where the Kronecker delta δ_i^k has the usual meaning.

The strain components in (3.14) are expressed in terms of components of the displacement vector with respect to B_0 and their covariant derivatives with respect to B_0. We may, of course, also express the strain-components in terms of components of the displacement vector with respect to B. For this purpose we write

$$\mathbf{r}=\mathbf{R}-\mathbf{v}+\mathbf{c}$$

and we obtain the components of the metric tensor of B_0 in the form

$$g_{ik}=G_{ik}-\mathbf{E}_i\cdot\mathbf{v}_{,k}-\mathbf{E}_k\cdot\mathbf{v}_{,i}+\mathbf{v}_{,i}\cdot\mathbf{v}_{,k}$$

and hence

$$\gamma_{ik}=\tfrac{1}{2}(\mathbf{E}_i\cdot\mathbf{v}_{,k}+\mathbf{E}_k\cdot\mathbf{v}_{,i}-\mathbf{v}_{,i}\cdot\mathbf{v}_{,k}). \quad . \quad . \quad . \quad (3.16)$$

The displacement vector, in terms of the contravariant base vectors $\mathbf{E}^i$ of the body B, is

$$\mathbf{v}=V_m\mathbf{E}^m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.17)$$

and its derivative is

$$\mathbf{v}_{,i}=V_m\|_i\mathbf{E}^m. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.18)$$

The double line denotes covariant differentiation with respect to B, that is with respect to θ_i and G_{ik}, so that

$$V_m\|_i=V_{m,i}-\begin{Bmatrix} r \\ mi \end{Bmatrix}V_r, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.19)$$

where the Christoffel symbols $\begin{Bmatrix} r \\ mi \end{Bmatrix}$ are calculated for B from G_{mn}, G^{mn} Thus

$$\left.\begin{aligned}\begin{Bmatrix} r \\ mi \end{Bmatrix}&=G^{rs}[s,mi],\\ [s,mi]&=\tfrac{1}{2}\{G_{si,m}+G_{sm,i}-G_{mi,s}\}.\end{aligned}\right\} \quad . \quad . \quad (3.20)$$

Since

$$\mathbf{E}_i\,.\,\mathbf{E}^k=\delta_i^k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.21)$$

we find from (3.16) and (3.18) that the strain components now become

$$\gamma_{ik}=\tfrac{1}{2}(V_i\|_k+V_k\|_i-V^r\|_iV_r\|_k), \quad . \quad . \quad . \quad . \quad . \quad (3.22)$$

which gives the strain-components in terms of the components of the displacement with respect to B.

The components of strain in terms of the displacement referred to any convenient set of coordinates in either B_0 or B may at once be written down from (3.14) and (3.22) respectively. For example, the strain-components referred to the cartesian system x_i in B_0 may be found from (3.14) by putting $\theta_i=x_i$, and are

$$\gamma_{11}=\frac{\partial v_1}{\partial x_1}+\frac{1}{2}\left\{\left(\frac{\partial v_1}{\partial x_1}\right)^2+\left(\frac{\partial v_2}{\partial x_1}\right)^2+\left(\frac{\partial v_3}{\partial x_1}\right)^2\right\},$$

$$\gamma_{23}=\frac{1}{2}\left(\frac{\partial v_2}{\partial x_3}+\frac{\partial v_3}{\partial x_2}\right)+\frac{1}{2}\left(\frac{\partial v_1}{\partial x_2}\frac{\partial v_1}{\partial x_3}+\frac{\partial v_2}{\partial x_2}\frac{\partial v_2}{\partial x_3}+\frac{\partial v_3}{\partial x_2}\frac{\partial v_3}{\partial x_3}\right),$$

etc. Similarly, when at time t the θ_i-curves coincide with the fixed cartesian y_i-curves, so that $\theta_i=y_i$, the strain components referred to y_i may be found from (3.22). Thus

$$\gamma_{11}=\frac{\partial V_1}{\partial y_1}-\frac{1}{2}\left\{\left(\frac{\partial V_1}{\partial y_1}\right)^2+\left(\frac{\partial V_2}{\partial y_1}\right)^2+\left(\frac{\partial V_3}{\partial y_1}\right)^2\right\};$$

$$\gamma_{23}=\frac{1}{2}\left(\frac{\partial V_2}{\partial y_3}+\frac{\partial V_3}{\partial y_2}\right)-\frac{1}{2}\left(\frac{\partial V_1}{\partial y_2}\frac{\partial V_1}{\partial y_3}+\frac{\partial V_2}{\partial y_2}\frac{\partial V_2}{\partial y_3}+\frac{\partial V_3}{\partial y_2}\frac{\partial V_3}{\partial y_3}\right),$$

etc.

The mixed components of the strain tensor can be formed in two ways according to the choice of metric. Using the strained metric components G^{ik} we have

$$\epsilon^i_k=G^{il}\gamma_{lk}, \qquad \text{(3.23)}$$

and using the components g^{ik} for the unstrained body,

$$\gamma^i_k=g^{il}\gamma_{lk}. \qquad \text{(3.24)}$$

Similarly, the contravariant components of the strain tensor may be formed in three ways. For our purpose the most convenient definition to take is

$$\epsilon^{ik}=g^{ir}\epsilon^k_r=G^{kl}\gamma^i_l=g^{ir}G^{kl}\gamma_{rl}. \qquad \text{(3.25)}$$

Using (3.2) this becomes

$$\epsilon^{ik}=\tfrac{1}{2}g^{ir}G^{kl}(G_{rl}-g_{rl}), \qquad \text{(3.26)}$$

which reduces to

$$\epsilon^{ik}=\tfrac{1}{2}(g^{ik}-G^{ik}) \qquad \text{(3.27)}$$

or

$$G^{ik}=g^{ik}-2\epsilon^{ik}, \qquad \text{(3.28)}$$

if we use the results

$$g^{ir}g_{rl}=\delta^i_l, \qquad G^{kl}G_{rl}=\delta^k_l. \qquad \text{(3.29)}$$

§ 4. Equations of Compatibility.

The components γ_{ik} of the strain tensor, as found in the preceding section, are not independent but have to satisfy the condition that the components of the Riemann–Christoffel tensors ${}_0R_{ijkl}$ and R_{ijkl} of the original and strained bodies must vanish. We may write the components of the Riemann–Christoffel tensor of the strained body in the form

$$R_{ijkl}=\tfrac{1}{2}(G_{il,jk}+G_{jk,il}-G_{ik,jl}-G_{jl,ik})$$
$$+G^{mn}\{[m,li][n,kj]-[m,ik][n,lj]\}. \qquad \text{(4.1)}$$

Using (3.3) the Cristoffel-symbols $[i, rs]$ of the strained medium become

$$[i,rs]=\tfrac{1}{2}(G_{ir,s}+G_{is,r}-G_{rs,i})=[i,rs]_0+2\gamma_{[i,rs]}, \qquad \text{(4.2)}$$

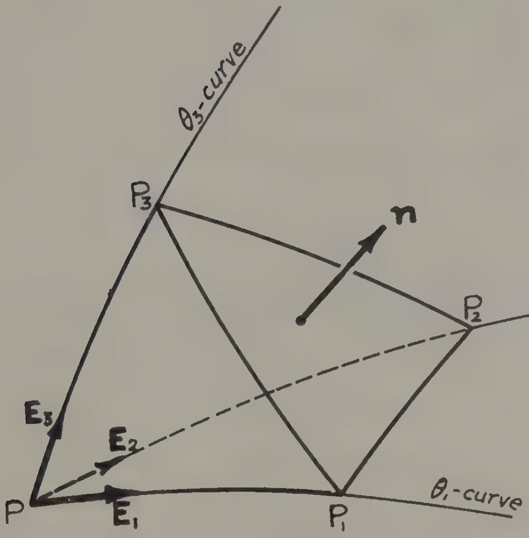
where we have put

$$\gamma_{[i,rs]}=\tfrac{1}{2}(\gamma_{ir,s}+\gamma_{is,r}-\gamma_{rs,i}). \qquad \text{(4.3)}$$

If we now substitute (3.28), (3.3) and (4.2) into (4.1) we obtain

$$\begin{aligned} R_{ijkl} = {}_0R_{ijkl} &+ \gamma_{il,jk} + \gamma_{jk,il} - \gamma_{ik,jl} - \gamma_{jl,ik} \\ &- 2\epsilon^{mn}([m, li]_0[n, kj]_0 - [m, ik]_0[n, lj]_0) \\ &+ 2g^{mn}\{[n, kj]_0\gamma_{[m,li]} + [m, li]_0\gamma_{[n,kj]} \\ &\qquad - [n, lj]_0\gamma_{[m,ik]} - [m, ik]_0\gamma_{[n,lj]}\} \\ &- 4\epsilon^{mn}\{[n, kj]_0\gamma_{[m,li]} + [m, li]_0\gamma_{[n,kj]} - [n, lj]_0\gamma_{[m,ik]} \\ &\qquad - [m, ik]_0\gamma_{[n,lj]} + 2\gamma_{[m,li]}\gamma_{[n,kj]} - 2\gamma_{[m,ik]}\gamma_{[n,lj]}\} \\ &+ 4g^{mn}(\gamma_{[m,li]}\gamma_{[n,kj]} - \gamma_{[m,ik]}\gamma_{[n,lj]}). \qquad \qquad (4.4) \end{aligned}$$

Fig. 2.



The conditions

$$ {}_0R_{ijkl} = 0, \qquad R_{ijkl} = 0, \qquad \qquad (4.5)$$

used in (4.4) give relations between the strain-components which are called the equations of compatibility. As we know that there are only six independent components of the Riemann–Christoffel tensor, (4.4) and (4.5) reduce to six independent equations.

§ 5. EXTERNAL FORCE. STRESS TENSOR.

We assume that the body B_0 is moved to its strained position B by the action of external forces which may vary with the time. At any point P of the strained body B we construct a parallelepiped which is bounded by the faces θ_i=constant, $\theta_i + d\theta^i$=constant, and a tetrahedron whose edges are formed by the coordinate curves PP of lengths ds_i, and the curves P_1P_2, P_2P_3, P_3P_1 (see fig. 2), which in the limit are defined by the three vectors

$$d\mathbf{s}_j - d\mathbf{s}_i = \mathbf{E}_j\,d\theta^j - \mathbf{E}_i\,d\theta^i \qquad (i, j \text{ not summed}). \quad . \quad . \quad . \quad (5.1)$$

The surfaces θ_i=constant of the tetrahedron have areas

$$\tfrac{1}{2}dS_i = \tfrac{1}{2}\sqrt{(GG^{ii})}d\theta^j\,d\theta^k \qquad (i \neq j \neq k,\ i \text{ not summed}), \quad . \quad (5.2)$$

and the area of $P_1P_2P_3$ is denoted by $\tfrac{1}{2}dS$.

If $\mathbf{n}$ is a unit vector normal to $P_1P_2P_3$ then

$$\mathbf{n}\,dS = (d\mathbf{s}_2 - d\mathbf{s}_1)\times(d\mathbf{s}_3 - d\mathbf{s}_1),$$

and, using (5.1), this becomes

$$\left.\begin{aligned}\mathbf{n}\,dS &= (\mathbf{E}_j\times\mathbf{E}_k)d\theta^j\,d\theta^k\\ &= \epsilon_{ijk}\mathbf{E}^i\,d\theta^j\,d\theta^k\end{aligned}\right\}(j, k = 1, 2;\ 2, 3;\ 3, 1), \quad . \quad . \quad (5.3)$$

where

$$\epsilon_{ijk}\mathbf{E}^i = \mathbf{E}_j\times\mathbf{E}_k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.4)$$

$$\left.\begin{aligned}\epsilon_{ijk} &= \sqrt{G} && \text{if } i, j, k \text{ is an even permutation of the numbers } 1, 2, 3,\\ &= -\sqrt{G} && \text{if } i, j, k \text{ is an odd permutation of the numbers } 1, 2, 3,\\ &= 0 && \text{if any two of the indices } i, j, k \text{ are equal and not summed.}\end{aligned}\right\} \quad . \quad . \quad . \quad (5.5)$$

Hence from (5.3) and (5.2)

$$\mathbf{n}\,dS = \sum_i \frac{\mathbf{E}^i}{\sqrt{G^{ii}}}dS_i. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.6)$$

If n_i are the covariant components of $\mathbf{n}$, with respect to the base vectors $\mathbf{E}^i$, it follows that †

$$n_i\sqrt{(G^{ii})}dS = dS_i. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.7)$$

A stress vector $\mathbf{t}$ is associated with the surface $P_1P_2P_3$ of the tetrahedron $PP_1P_2P_3$. The stress associated with the surfaces θ_i=constant of the tetrahedron can be represented in magnitude and direction by quantities $-\mathbf{t}_i$. The quantities $-\mathbf{t}_i$ are not invariant under general transformations of coordinates. This fact is apparent from the notation since vector invariants do not contain suffixes. The suffixes, however, do not indicate that the quantity transforms according to the covariant type of transformation.

By considering the equation of motion of the infinitesimal tetrahedron $PP_1P_2P_3$ we have, in the limit,

$$\mathbf{t}\,dS = \mathbf{t}_i\,dS_i. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.8)$$

---

† We remember that the summation convention is suspended when an index appears more than twice.

Volume forces and mass-accelerations acting on the tetrahedron do not appear in this equation since they are of a higher order of smallness than the surface forces.

From (5.7) and (5.8) we find that

$$\mathbf{t}=\sum_i n_i \mathbf{t}_i \sqrt{G^{ii}}. \qquad (5.9)$$

Since, under general transformations of coordinates, $\mathbf{t}$ is an invariant and n_i are the covariant components of a vector, it follows from (5.9) that $\mathbf{t}_i\sqrt{G^{ii}}$ transforms according to a contravariant type of transformation. We may therefore write

$$\mathbf{t}_i\sqrt{G^{ii}}=\tau^{ik}\mathbf{E}_k=\tau^i_k\mathbf{E}^k, \qquad (5.10)$$

where τ^{ik} and τ^i_k are the contravariant and mixed components of a tensor of the second order which is called the *stress tensor*. Later we will prove that the contravariant components are symmetric so that we are allowed to write the mixed components as τ^i_k and not $\tau^{i}{}_{.k}$ or $\tau_{k}{}^{i}$.

The stress quantities $\mathbf{t}_i$ exerted by the body in the region $\theta_i>0$ on the body in the region $\theta_i<0$, across the θ_i-surface, has positive values for its components when τ^{ik} are positive.

The covariant components of the stress tensor may be defined by the equations

$$\tau_{ik}=G_{ir}\tau^r_k. \qquad (5.11)$$

From (5.10) we see that the contravariant components τ^{ik} of the stress tensor and the mixed components τ^i_k of the stress tensor are related to the quantities $\mathbf{t}_i\sqrt{G^{ii}}$. The covariant components τ_{ik} of the stress tensor cannot be related to the quantities $\mathbf{t}_i\sqrt{G^{ii}}$ in a simple way and are therefore, in general, of less importance.

From (5.9) and (5.10) we obtain

$$\mathbf{t}=\tau^{ik}n_i\mathbf{E}_k=\tau^i_k n_i\mathbf{E}^k, \qquad (5.12)$$

so that if

$$\mathbf{t}=t^k\mathbf{E}_k=t_k\mathbf{E}^k, \qquad (5.13)$$

then

$$t^k=\tau^{ik}n_i, \qquad t_k=\tau^i_k n_i, \qquad (5.14)$$

and t^k, t_k are the contravariant and covariant components respectively of the stress vector $\mathbf{t}$.

The three stress quantities $\mathbf{t}_i$ in (5.10) may be put in the form

$$\mathbf{t}_i=\sum_k \sqrt{\left(\frac{G_{kk}}{G^{ii}}\right)}\,\tau^{ik}\frac{\mathbf{E}_k}{\sqrt{G_{kk}}}, \qquad (5.15)$$

where $\mathbf{E}_k/\sqrt{G_{kk}}$ are unit vectors along the coordinate curves. The *physical* stress quantities are therefore

$$\sqrt{\left(\frac{G_{kk}}{G^{ii}}\right)}\,\tau^{ik}. \qquad (5.16)$$

These quantities are not, of course, the components of a tensor.

§ 6. Equations of Motion and Equilibrium.

We consider the infinitesimal parallelepiped in fig. 3 whose faces have areas dS_i and we consider the equations of motion of all the forces acting on this elementary volume. The forces on the faces θ_i=constant through P are, to the first order, $-\mathbf{t}_i\, dS_i$ (i not summed), and since

$$dS_i=\sqrt{(GG^{ii})}\,d\theta^j\, d\theta^k \qquad (i \text{ not summed, } i \neq j \neq k), \qquad (6.1)$$

we may write this as

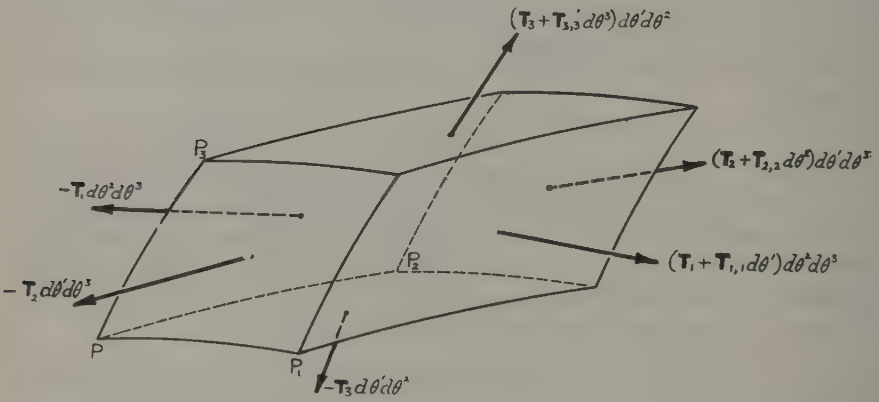
$$-\mathbf{T}_i\, d\theta^j\, d\theta^k \qquad (i \neq j \neq k),$$

where

$$\mathbf{T}_i=\sqrt{(GG^{ii})}\,\mathbf{t}_i=\sqrt{(G)}\,\tau^{ik}\mathbf{E}_k=\sqrt{(G)}\,\tau^i_k\mathbf{E}^k. \quad . \quad . \quad . \quad . \quad (6.2)$$

The forces on the faces θ_i=constant and on the faces $\theta_i+d\theta^i$=constant are shown in fig. 3.

Fig. 3.


$(\mathbf{T}_3+\mathbf{T}_{3,3}\,d\theta^3)\,d\theta^1 d\theta^2$
P_3
$-\mathbf{T}_1 d\theta^2 d\theta^3$
$(\mathbf{T}_2+\mathbf{T}_{2,2}\,d\theta^2)\,d\theta^1 d\theta^3$
P_2
$(\mathbf{T}_1+\mathbf{T}_{1,1}\,d\theta^1)\,d\theta^2 d\theta^3$
$-\mathbf{T}_2 d\theta^1 d\theta^3$
P
P_1
$-\mathbf{T}_3 d\theta^1 d\theta^2$

The body forces acting on the elementary parallelepiped whose volume is $d\tau$ are $\rho\mathbf{F}\, d\tau$ or $\rho\mathbf{Q}\, d\theta^1\, d\theta^2\, d\theta^3$, where

$$\mathbf{Q}=\mathbf{F}\sqrt{G}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.3)$$

and where ρ is the density of the strained body B.

We now take moments about the centre of gravity of the elementary volume, and, since in the limit we may retain only first order terms, this gives

$$\mathbf{E}_i\times\mathbf{T}_i=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.4)$$

From (6.2), (5.4) and (6.4) we obtain

$$\epsilon_{ikl}\tau^{ik}\mathbf{E}^l=0.$$

It follows immediately from this equation that

$$\tau^{ik}=\tau^{ki}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.5)$$

so that the contravariant components (and therefore also the covariant components) of the stress tensor are symmetric.

Since the components τ^{ik} are symmetric it is always possible to find a coordinate system in B in which $\tau^{ik}=0$ $(i \neq k)$. The directions of the coordinate curves in such a system are called the principal axes of stress and the stresses which correspond to these axes are called the *principal stresses.*

The remaining equation of motion states that the (vector) sum of all the forces acting on the elementary volume equals the mass-acceleration of the volume, where the acceleration may be taken to be that of the centre of gravity. Hence, in the limit, (see fig. 3)

$$\mathbf{T}_{i,\,i}+\rho\mathbf{Q}=\rho\mathbf{h}, \qquad (6.6)$$

where

$$\mathbf{h}=\mathbf{f}\sqrt{G}, \qquad (6.7)$$

and **f** is the acceleration vector.

When the body is in equilibrium $\mathbf{f}=0$ and

$$\mathbf{T}_{i,\,i}+\rho\mathbf{Q}=0. \qquad (6.8)$$

We may express the equation of motion in alternative forms. If

$$\left.\begin{aligned}\mathbf{F}&=F^i\mathbf{E}_i=F_i\mathbf{E}^i,\\ \mathbf{f}&=f^i\mathbf{E}_i=f_i\mathbf{E}^i,\end{aligned}\right\} \qquad (6.9)$$

then

$$\left.\begin{aligned}\mathbf{Q}&=Q^i\mathbf{E}_i=Q_i\mathbf{E}^i,\\ \mathbf{h}&=h^i\mathbf{E}_i=h_i\mathbf{E}^i,\end{aligned}\right\} \qquad (6.10)$$

where

$$\left.\begin{aligned}Q^i&=F^i\sqrt{G}, \qquad Q_i=F_i\sqrt{G},\\ h^i&=f^i\sqrt{G}, \qquad h_i=f_i\sqrt{G}.\end{aligned}\right\} \qquad (6.11)$$

Also, if

$$\sigma^{ik}=\tau^{ik}\sqrt{G}, \qquad \sigma^i_k=\tau^i_k\sqrt{G}, \qquad (6.12)$$

then, from (6.2),

$$\mathbf{T}_i=\sigma^{ik}\mathbf{E}_k=\sigma^i_k\mathbf{E}^k. \qquad (6.13)$$

It should be noticed that the quantities defined in (6.11) and (6.12) are not components of tensors†. Using (6.9) to (6.13) the equation of motion (6.6) becomes

$$\sigma^{ik}{}_{,i}+\begin{Bmatrix}k\\ ir\end{Bmatrix}\sigma^{ir}+\rho Q^k=\rho h^k, \qquad (6.14)$$

or

$$\sigma^i_{k,\,i}-\begin{Bmatrix}r\\ ki\end{Bmatrix}\sigma^i_r+\rho Q_k=\rho h_k. \qquad (6.15)$$

Since

$$\begin{Bmatrix}i\\ ik\end{Bmatrix}=\frac{1}{\sqrt{G}}\frac{\partial\sqrt{G}}{\partial\theta_k},$$

---

† They are sometimes called relative or weighted tensors.

these equations may be reduced respectively to

$$\tau^{ik}{}_{,i}+\begin{Bmatrix}k\\ ir\end{Bmatrix}\tau^{ir}+\begin{Bmatrix}r\\ ir\end{Bmatrix}\tau^{ik}+\rho F^k=\rho f^k, \quad . \quad . \quad . \quad . \quad (6.16)$$

$$\tau^i_{k,i}-\begin{Bmatrix}r\\ ki\end{Bmatrix}\tau^i_r+\begin{Bmatrix}r\\ ir\end{Bmatrix}\tau^i_k+\rho F_k=\rho f_k. \quad . \quad . \quad . \quad . \quad (6.17)$$

These may be written more compactly in the tensor forms

$$\tau^{ik}\|_i+\rho F^k=\rho f^k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.18)$$

$$\tau^i_k\|_i+\rho F_k=\rho f_k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.19)$$

where the double line again denotes covariant differentiation with respect to the strained body B.

With the help of (5.11) it is possible, of course, to write the equations of motion in terms of the covariant components of the stress tensor but this form appears to be of little interest and will not be given here.

The equations (6.6) and (6.14) to (6.19) are alternative forms of the equations of motion. The most convenient form to use depends on the type of problem which is being examined but the equations (6.6), (6.18) and (6.19) are particularly simple in form.

Finally, the conditions at the boundary surface of the body at which the surface forces are prescribed require that,

$$\mathbf{t}=\mathbf{P}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.20)$$

where **P** is the surface force vector. If we express the surface forces **P** in terms of their components,

$$\mathbf{P}=\mathrm{P}^k\mathbf{E}_k=\mathrm{P}_k\mathbf{E}^k \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.21)$$

and use (5.12), the boundary conditions have the alternative forms

$$\tau^{ik}n_i=\mathrm{P}^k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.22)$$

or

$$\tau^i_k n_i=\mathrm{P}_k. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.23)$$

§ 7. VIRTUAL WORK.

When the body is in the strained state B at time t the displacement **v** of each point P which was originally at P_0 is

$$\mathbf{v}=\mathbf{v}(\theta_1, \theta_2, \theta_3, t).$$

We now consider any other geometrically possible motion which is such that, during the same time t, the points P_0 move to P′ and form a strained body B′, such that the displacement of P_0 is of the form

$$\mathbf{v}'=\mathbf{v}(\theta_1, \theta_2, \theta_3, t)+\epsilon\mathbf{w}(\theta_1, \theta_2, \theta_3, t),$$

where $\mathbf{w}$ is an arbitrary vector and ϵ is a real number which we shall take to be small, so that its squares and higher powers can be neglected compared with ϵ. We call $\epsilon\mathbf{w}$ a variation or virtual displacement of $\mathbf{v}$ and write

$$\delta\mathbf{v}=\mathbf{v}'-\mathbf{v}=\epsilon\mathbf{w}.$$

This corresponds to the usual definition of variation in the calculus of variations†.

The variation of $\mathbf{v}_{,i}$ is defined to be

$$\delta(\mathbf{v}_{,i})=\mathbf{v}'_{,i}-\mathbf{v}_{,i}=\epsilon\mathbf{w}_{,i}=(\epsilon\mathbf{w})_{,i}=(\delta\mathbf{v})_{,i}, \quad . \quad . \quad . \quad (7.1)$$

so that δ and partial derivatives with respect to θ_i are interchangeable.

The variations of the covariant base vectors $\mathbf{E}_i$ are given by

$$\delta\mathbf{E}_i=\delta(\mathbf{e}_i+\mathbf{v}_{,i})=(\delta\mathbf{v})_{,i},$$

if we remember that $\mathbf{e}_i$ does not depend on $\mathbf{v}$. The variations of the scalar-products of the base vectors are found to be

$$\delta(\mathbf{E}_i\,.\,\mathbf{E}_k)=\mathbf{E}_i\,.\,\delta\mathbf{E}_k+\mathbf{E}_k\,.\,\delta\mathbf{E}_i.$$

Hence, using (7.1) and (3.2), the variation of the components of the strain tensor are

$$2\delta\gamma_{ik}=\delta G_{ik}=\mathbf{E}_i\,.\,(\delta\mathbf{v})_{,k}+\mathbf{E}_k\,.\,(\delta\mathbf{v})_{,i}, \quad . \quad . \quad . \quad . \quad (7.2)$$

since $\delta g_{ik}=0$.

We now consider the whole strained body and calculate the work done, during a virtual displacement, by the body forces and the surface forces acting on its boundary. This is called the virtual work and is defined ‡ to be

$$\delta^*A=\int_S \mathbf{P}\,.\,\delta\mathbf{v}\,dS+\int_\tau \rho\mathbf{F}\,.\,\delta\mathbf{v}\,d\tau, \quad . \quad . \quad . \quad . \quad (7.3)$$

where dS is an element of the bounding surface and $d\tau$ a volume element. In (7.3) the surface integral is to be taken over the whole of the bounding surface S. The volume integral is throughout the volume of B. In order to be able to evaluate the surface integral in (7.3) we assume that the surface forces are prescribed over parts of the surface and that the surface displacements are prescribed over the remainder of the surface. If the variation of $\mathbf{v}$ is then chosen to be zero where the surface displacements are prescribed, the surface integral is zero over these parts and may be evaluated over the remainder of the surface over which the surface forces are known.

Using the boundary conditions (6.20) the expression for δ^*A may be written

$$\delta^*A=\int_S \mathbf{t}\,.\,\delta\mathbf{v}\,dS+\int_\tau \rho\mathbf{F}\,.\,\delta\mathbf{v}\,d\tau. \quad . \quad . \quad . \quad . \quad (7.4)$$

---

† See, for example, Courant and Hilbert (1931).

‡ The star on δ^* indicates that δ^*A is not the variation of a function A.

From (5.12) and (6.2)

$$\mathbf{t}=\frac{n_i}{\sqrt{G}}\mathbf{T}_i, \qquad \text{(7.5)}$$

so that (7.4) becomes

$$\delta^*A=\int_S \frac{n_i}{\sqrt{G}}\mathbf{T}_i \,.\, \delta\mathbf{v}\,dS+\int_\tau \rho\mathbf{F}\,.\,\delta\mathbf{v}\,d\tau.$$

The surface integral may be transformed by using Green's theorem in the form

$$\int_S B^r n_r\,dS=\int_\tau \frac{\{B\sqrt{G}\}_{,r}}{\sqrt{G}}d\tau, \qquad \text{(7.6)}$$

where B^r are the contravariant components of a vector. Thus

$$\delta^*A=\int_\tau \frac{(\mathbf{T}_i\,.\,\delta\mathbf{v})_{,i}}{\sqrt{G}}d\tau+\int_\tau \rho\mathbf{F}\,.\,\delta\mathbf{v}\,d\tau=\int_\tau \{(\mathbf{T}_{i,i}+\rho\mathbf{Q})\delta\mathbf{v}+\mathbf{T}_i\,.\,(\delta\mathbf{v})_{,i}\}\frac{d\tau}{\sqrt{G}},$$

and using (6.2) and (6.6) this reduces to

$$\delta^*A=\int_\tau \{\rho\mathbf{f}\,.\,\delta\mathbf{v}+\tau^{ik}\mathbf{E}_k\,.\,(\delta\mathbf{v})_{,i}\}\,d\tau. \qquad \text{(7.7)}$$

Since τ^{ik} are symmetric components of a tensor

$$\delta^*A=\int_\tau \{\rho\mathbf{f}\,.\,\delta\mathbf{v}+\tfrac{1}{2}\tau^{ik}[\mathbf{E}_k\,.\,(\delta\mathbf{v})_{,i}+\mathbf{E}_i\,.\,(\delta\mathbf{v})_{,k}]\}\,d\tau$$

$$=\iiint \{\rho\mathbf{f}\,.\,\delta\mathbf{v}+\tau^{ik}\delta\gamma_{ik}\}\sqrt{G}\,d\theta^1\,d\theta^2\,d\theta^3, \qquad \text{(7.8)}$$

if we use (7.2).

Subject to boundary conditions at the surface of the body, the virtual displacement is an arbitrary geometrically possible displacement, so that equation (7.8), as is seen from its derivation, can replace the equations of motion of the body.

When the body is in equilibrium, $\mathbf{f}=0$, and (7.8) reduces to †

$$\delta^*A=\iiint \tau^{ik}(\delta\gamma_{ik})\sqrt{G}\,d\theta^1\,d\theta^2\,d\theta^3. \qquad \text{(7.9)}$$

This represents the necessary (and sufficient) conditions that the body is in equilibrium and may replace the equations of equilibrium.

For a rigid virtual displacement $\delta\gamma_{ik}=0$ and therefore

$$\delta^*A=0. \qquad \text{(7.10)}$$

This states that for a body in equilibrium the virtual work of the external forces is zero for any rigid virtual displacements.

---

† This result is a general form of the well-known principle of virtual work of elastic bodies.

§ 8. STRAIN-ENERGY FUNCTION.

We now confine our attention to a body B_0 which at time $t=0$ is unstrained and unstressed †, and we assume that B is obtained from B_0 entirely by the action of external forces ‡. If B is able to recover its original size and shape *completely* when the external forces cease to act the body is said to be elastic. This implies that when all the components of the strain tensor vanish then all the components of the stress tensor vanish and vice versa.

An element of mass of the strained body B at time t is given by

$$dm=\rho\sqrt{G}\,d\theta^1\,d\theta^2\,d\theta^3, \quad . \quad . \quad . \quad . \quad . \quad (8.1)$$

and the principle of conservation of mass in a virtual displacement is expressed by

$$\delta(dm)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.2)$$

We now consider the actual motion of the body B which takes place in time dt so that the displacement vector increases by $d\mathbf{v}=\frac{\partial \mathbf{v}}{\partial t}$, θ_i being fixed. Let dK denote the increment of kinetic energy of B during the time dt. The kinetic energy K is given by

$$K=\tfrac{1}{2}\int_B\left(\frac{\partial \mathbf{v}}{\partial t}\right)^2 dm, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.3)$$

so that

$$dK=\int\frac{\partial \mathbf{v}}{\partial t}\,.\,d\left(\frac{\partial \mathbf{v}}{\partial t}\right)dm,$$

since the mass element dm is unaltered. But $d\left(\frac{\partial \mathbf{v}}{\partial t}\right)$ is the increase of velocity in time dt and therefore

$$d\left(\frac{\partial \mathbf{v}}{\partial t}\right)=\frac{\partial^2 \mathbf{v}}{\partial t^2}\,dt=\mathbf{f}\,dt.$$

Hence

$$dK=\int_B \mathbf{f}\,.\,d\mathbf{v}\,dm=\iiint(\mathbf{f}\,.\,d\mathbf{v})\rho\sqrt{G}\,d\theta^1\,d\theta^2\,d\theta^3. \quad . \quad . \quad (8.4)$$

The actual increase in the work done d^*A by the external forces acting on B, during time dt, may be obtained by putting $\delta\mathbf{v}=d\mathbf{v}$ and

$$\delta\gamma_{ik}=d\gamma_{ik}=\frac{\partial\gamma_{ik}}{\partial t}\,dt,$$

---

† This means, for example, that no stresses are induced in B_0 by its manufacture or by the action of body forces.

‡ Strains produced without external forces but by changes of temperature are not considered.

in (7.8), where $d\gamma_{ik}$ are the increases in the components of strain in time dt. Thus

$$d^*\mathrm{A}=\iiint(\rho\mathbf{f}\,.\,d\mathbf{v}+\tau^{ik}\,d\gamma_{ik})\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3. \quad . \quad . \quad . \quad (8.5)$$

The fundamental energy-conservation laws of thermodynamics for reversible processes may be written†

$$\int_{\mathrm{B}}\mathrm{T}\,d\sigma\,dm+d^*\mathrm{A}=\int_{\mathrm{B}}d\mathrm{U}\,dm+d\mathrm{K}, \quad . \quad . \quad . \quad . \quad (8.6)$$

where T is the temperature of the mass element dm, $d\sigma$ the increment of entropy per unit mass and $d\mathrm{U}$ the increment of internal energy per unit mass. By using (8.4) and (8.5) this equation reduces to

$$\iiint(d\mathrm{U}-\mathrm{T}\,d\sigma)\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3=\iiint\tau^{ik}(d\gamma_{ik})\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3. \quad (8.7)$$

Since we may consider any arbitrary volume, it follows that

$$(d\mathrm{U}-\mathrm{T}\,d\sigma)\rho=\tau^{ik}\,d\gamma_{ik}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.8)$$

if we make suitable assumptions about the continuity of the integrands.

If the changes which have taken place in time dt are reversible and isothermal, then T is constant, and

$$d\mathrm{W}=d(\mathrm{U}-\mathrm{T}\sigma)=\frac{\tau^{ik}}{\rho}\,d\gamma_{ik}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.9)$$

where $d\mathrm{W}$ is a complete differential. The quantities $d\gamma_{ik}$ are also, for fixed values of θ_i, complete differentials. Since we are assuming that the body is elastic, the changes which take place are reversible, and the condition of the body B is independent of the actual path taken to reach the state B from B_0. The differentials $d\gamma_{ik}$ in (8.9) may therefore be considered as arbitrary so that, since $d\mathrm{W}$ is a complete differential,

$$\tau^{ik}=\rho\frac{\partial\mathrm{W}}{\partial\gamma_{ik}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.10)$$

where W is a function of γ_{ik} and all changes of state are isothermal.

Similarly, if the changes in an elastic body are adiabatic, so that $d\sigma=0$, we find from (8.8)

$$\tau^{ik}=\rho\frac{\partial\mathrm{U}}{\partial\gamma_{ik}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.11)$$

where U is a function of γ_{ik}.

The function W and U are called strain-energy functions and play the rôle of potential energy per unit mass for the isothermal and adiabatic cases respectively.

---

† Under these conditions $d\mathrm{U}$ and $d\sigma$ are complete differentials.

The relations (8.10) and (8.11) between the contravariant components of the stress tensor and the strain-energy function is the most convenient for our purpose, but it is of interest to derive an alternative form†. The mixed components ϵ^i_k of the strain tensor are defined in (3.23) and from this definition

$$\gamma_{kl} = \epsilon^i_l G_{ki} = \epsilon^i_l (g_{ki} + 2\gamma_{ki})$$

Hence

$$d\gamma_{kl} = G_{ki}\, d\epsilon^i_l + 2\epsilon^i_l\, d\gamma_{ki}$$

and therefore

$$G^{mk}(\delta^i_l - 2\epsilon^i_l) d\gamma_{ki} = G^{mk} G_{ki}\, d\epsilon^i_l = d\epsilon^m_l. \quad . \quad . \quad . \quad . \quad (8.12)$$

From (8.10)

$$\rho\, dW = \tau^{ik}\, d\gamma_{ik} = G^{mk} \tau^i_m\, d\gamma_{ki},$$

and with (8.12) this becomes

$$\rho(\delta^i_l - 2\epsilon^i_l)\, dW = \tau^i_m\, d\epsilon^m_l,$$

so that ‡, for the reversible isothermal case,

$$\tau^i_m = \rho\left(\frac{\partial W}{\partial \epsilon^m_i} - 2\epsilon^i_l \frac{\partial W}{\partial \epsilon^m_l}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.13)$$

with a similar result with U written for W when the changes are reversible and adiabatic.

§ 9. VARIATIONAL, WORK AND ENERGY PRINCIPLES.

In this, and in subsequent sections, we shall consider an elastic body which undergoes isothermal changes of state and use the strain-energy function W. The corresponding results for adiabatic changes may be obtained in a similar manner.

If W′ is the total strain-energy of the whole body, then

$$W' = \iiint \rho W \sqrt{G}\, d\theta^1\, d\theta^2\, d\theta^3, \quad . \quad . \quad . \quad . \quad . \quad (9.1)$$

and, remembering § (8.2),

$$\begin{aligned} \delta W' &= \iiint (\delta W) \rho \sqrt{G}\, d\theta^1\, d\theta^2\, d\theta^3 \\ &= \iiint \frac{\partial W}{\partial \gamma_{ik}} (\delta \gamma_{ik}) \rho \sqrt{G}\, d\theta^1\, d\theta^2\, d\theta^3. \quad . \quad . \quad . \quad (9.2) \end{aligned}$$

---

† Other relations between the components of the stress tensor and the strain energy may be found by using the expression (7.7) for δ^*A instead of (7.8).

‡ If our coordinates θ_i are identified with a general coordinate system in the strained body B at a given time t, then (8.13) agrees with a similar result obtained by a somewhat different argument by F. D. Murnaghan (1937). On the other hand, if we take θ_i to be a general coordinate system in the original body B_0, the result (8.10) is the same as that given by Murnaghan when his two coordinate systems coincide.

§ The δ used here is the usual variation symbol of the Calculus of Variations.

Hence, from (7.8), (8.10) and (9.2), we have

$$\delta^*\mathrm{A}=\delta\mathrm{W}'+\iiint(\mathbf{f}\,.\,\delta\mathbf{v})\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3. \quad . \quad . \quad . \quad (9.3)$$

We integrate this last result between two arbitary times t_0 and t_1. Thus

$$\int_{t_0}^{t_1}\delta^*\mathrm{A}\,dt=\int_{t_0}^{t_1}\delta\mathrm{W}'\,dt+\int_{t_0}^{t_1}\iiint(\mathbf{f}\,.\,\delta\mathbf{v})\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3\,dt. \quad . \quad (9.4)$$

The last integral may be transformed by partial integration. Using $\mathbf{f}=\dfrac{\partial^2\mathbf{v}}{\partial t^2}$, and remembering that the mass element $\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3$ is constant, we obtain

$$\int_{t_0}^{t_1}\iiint(\mathbf{f}\,.\,\delta\mathbf{v})\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3\,dt=\left[\iiint\left(\frac{\partial\mathbf{v}}{\partial t}\,.\,\delta\mathbf{v}\right)\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3\right]_{t_0}^{t_1}$$
$$-\int_{t_0}^{t_1}\iiint\frac{\partial\mathbf{v}}{\partial t}\,.\,\frac{\partial\delta\mathbf{v}}{\partial t}\,\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3\,dt.$$

We now suppose that the variation $\delta\mathbf{v}$ is such that it vanishes at time t_0 and t_1 so that

$$\int_{t_0}^{t_1}\iiint(\mathbf{f}\,.\,\delta\mathbf{v})\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3\,dt=-\int_{t_0}^{t_1}\delta\mathrm{K}\,dt, \quad . \quad . \quad (9.5)$$

where K, the kinetic energy of the strained body B, is given by (8.3), and

$$\delta\mathrm{K}=\iiint\frac{\partial\mathbf{v}}{\partial t}\,.\,\delta\left(\frac{\partial\mathbf{v}}{\partial t}\right)\rho\sqrt{\mathrm{G}}\,d\theta^1\,d\theta^2\,d\theta^3. \quad . \quad . \quad . \quad . \quad (9.6)$$

Equation (9.4) now becomes

$$\int_{t_0}^{t_1}\delta^*\mathrm{A}\,dt=\int_{t_0}^{t_1}\delta\mathrm{W}'\,dt-\int_{t_0}^{t_1}\delta\mathrm{K}\,dt. \quad . \quad . \quad . \quad . \quad (9.7)$$

If the external forces can be derived from a potential energy function Ω then

$$\delta^*\mathrm{A}=-\delta\Omega, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.8)$$

and (9.7) reduces to the simple form

$$\int_{t_0}^{t_1}\delta(\mathrm{K}-\mathrm{W}'-\Omega)dt=\delta\int_{t_0}^{t_1}(\mathrm{K}-\mathrm{W}'-\Omega)dt=0, \quad . \quad . \quad . \quad (9.9)$$

where $\mathrm{K}-\mathrm{W}'-\Omega$ is the Lagrangian function.

Equation (9.9) is the analytical expression of Hamilton's Principle, which says that the variation of the integral of the Lagrangian function, between times t_0 and t_1, takes a stationary value, provided that the variation of the displacement vector is taken in such a way that it vanishes at the times t_0, t_1.

If the strained body B is in equilibrium then $\mathbf{f}=0$ and the virtual work (9.3) of the external forces acting on the body becomes

$$\delta^*\mathrm{A}=\delta\mathrm{W}',$$

and, if (9.8) still holds,

$$\delta(\mathrm{W}'+\Omega)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.10)$$

This states that the variation of the total potential energy $W'+\Omega$ has a stationary value.

The principles (9.9) and (9.10) were found from the expression (9.3) for the virtual work of the external forces. We may also derive from (9.3) expressions for the *actual* work performed in time dt by replacing $\delta\mathbf{v}$ by the actual displacement $d\mathbf{v}$. The virtual work of the external forces δ^*A now becomes the actual increase d^*A of work performed in time dt, and $\delta W'$ becomes dW', and therefore (9.3) is now

$$d^*A=dW'+\iiint(\mathbf{f}\,.\,d\mathbf{v})\rho\sqrt{G}\,d\theta^1\,d\theta^2\,d\theta^3. \qquad . \quad . \quad . \quad (9.11)$$

If we use (8.4) this reduces to

$$d^*A=dW'+dK. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.12)$$

This states that the work done by the external forces in time dt is equal to the sum of the increase of the kinetic energy and the increase of the strain-energy.

§ 10. STRESS-STRAIN RELATIONS.

We have seen in § 8 that, under certain conditions, the stresses which occur in an elastic body are related to the function W by the equations

$$\tau^{ik}=\rho\frac{\partial W}{\partial\gamma_{ik}}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10.1)$$

and

$$W'=\iiint\rho W\sqrt{G}\,d\theta^1\,d\theta^2\,d\theta^3, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (10.2)$$

represents the total strain-energy of the strained body. Since mass elements are conserved under straining, we have

$$\rho_0\sqrt{g}=\rho\sqrt{G}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10.3)$$

where ρ_0 is the density of the unstrained body. Hence equations (10.1) may be written

$$\sqrt{(G)}\tau^{ik}=\rho_0\sqrt{(g)}\frac{\partial W}{\partial\gamma_{ik}}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (10.4)$$

At this point it is necessary to make further assumptions if we wish to obtain an explicit relation between stress and strain. The validity of the assumptions when applied to physical bodies must be tested by referring to experiment. The relation between stress and strain, in the theory which has been developed so far, is of such a form that a strain-energy function W exists with the property (10.4). We therefore assume a suitable form for W and derive the stress-strain relations from (10.4).

Since the density ρ_0 of the unstrained body, and W, are scalars, the quantity $\rho_0 W$ is invariant under all transformations of coordinates and we assume that W can be expanded as a power series[†] in the strain components γ_{ik}. Thus

$$\rho_0 W = E + E^{ik}\gamma_{ik} + F^{iklr}\gamma_{ik}\gamma_{lr} + F^{iklrmn}\gamma_{ik}\gamma_{lr}\gamma_{mn} + \ldots, \qquad (10.5)$$

The functions E^{ik}, F^{iklr}, ... do not depend on γ_{ik} and since W is invariant they must be components of tensors. If W is taken to be zero when the body is in the initial state B in which $\gamma_{ik}=0$ then $E=0$.

From (10.4) and (10.5) we have[‡]

$$\sqrt{(G/g)}\,\tau^{ik} = E^{ik} + E^{iklr}\gamma_{lr} + E^{iklrmn}\gamma_{lr}\gamma_{mn} + \ldots, \qquad (10.6)$$

where we have put

$$\left.\begin{aligned} E^{iklr} &= F^{iklr} + F^{lrik}, \\ E^{iklrmn} &= F^{iklrmn} + F^{lrikmn} + F^{mnlrik}. \end{aligned}\right\} \qquad (10.7)$$

Since we have assumed that our material is elastic there is no stress when the strains are zero provided there is no initial strain. Hence

$$E^{ik} = 0.$$

The expression (10.7) gives us a general relation between stress and strain, and we call the coefficients E^{iklr}, E^{iklrmn}, ... the elastic coefficients of the body. They depend on the properties of the body in the unstrained state.

The quantities E^{iklr} are coefficients in an expansion of the strain-energy function W in terms of the components of strain and therefore refer to an elastic body which undergoes isothermal changes of state. When the changes are adiabatic the strain-energy function is U and the corresponding elastic coefficients are, in general, different, but for many purposes these differences are not great and can be ignored.

From the symmetric properties of the components of the stress and strain tensors, and the relations (10.7), we deduce the following symmetric properties :

$$E^{iklr} = E^{kilr} = E^{ikrl} = E^{lrik}, \qquad (10.8)$$

with similar properties for E^{iklrmn}, ...

Because of the symmetric properties of the elastic coefficients, we see, from (10.4) and (10.6), that the strain-energy (10.5) may be expressed in the more convenient form

$$\rho_0 W = \tfrac{1}{2}E^{iklr}\gamma_{ik}\gamma_{lr} + \tfrac{1}{3}E^{iklrmn}\gamma_{ik}\gamma_{lr}\gamma_{mn}. \qquad (10.9)$$

In order to examine the properties of the elastic coefficients in more detail we refer all quantities to the rectangular coordinates x_i in the unstrained body B_0 so that $\theta_i = x_i$ and $g=1$. In these coordinates the tensor components E^{iklr}, E^{iklrmn}, ... are denoted by c^{iklr}, c^{iklrmn}, and the covariant components γ_{mn} of the strain tensor by e_{mn}. The contravariant components of the stress tensor are denoted by t^{ik} and these are referred to coordinates in the strained body B which, in the unstrained body B_0, were the coordinates x_i. The strain-energy function then becomes

---

[†] This, of course, is not the only possible assumption that can be made.

[‡] In using (10.4) we suppose that γ_{ik} is distinct from γ_{ki}.

$$\rho_0 W = \tfrac{1}{2}c^{iklr}e_{ik}e_{lr} + \tfrac{1}{3}c^{iklrmn}e_{ik}e_{lr}e_{mn} + \dots \tag{10.10}$$

and

$$t^{ik}\sqrt{(G)} = c^{iklr}e_{lr} + c^{iklrmn}e_{lr}e_{mn} + \dots \tag{10.11}$$

If the elastic coefficients c^{iklr}, $c^{iklrmn}, \dots$ are independent of the orientation of the rectangular axes x_i at each point we say that the body is *elastically isotropic*, or more simply, *isotropic*. Otherwise the body is said to be æolotropic or anisotropic.

If the density of the original body B_0 is constant and if the elastic coefficients c^{iklr}, $c^{iklrmn}, \dots$ are constants, we say that the body is *elastically homogeneous*. In this case the covariant derivatives of the general elastic coefficients E^{iklr}, $E^{iklrmn}, \dots$ are zero, so that

$$E^{iklr}|_t = 0, \qquad E^{iklrmn}|_t = 0. \tag{10.12}$$

In the remainder of this paper we shall restrict our attention to bodies which are homogeneous in the above sense so that equations (10.12) hold. Similar results, however, hold for non-homogeneous bodies. Because of theoretical and experimental difficulties we restrict our attention to those materials for which the coefficients $E^{iklrmn}, \dots$ are sufficiently small, so that we can express the strain-energy W approximately in the form

$$\rho_0 W = \tfrac{1}{2}E^{iklr}\gamma_{ik}\gamma_{lr} = \tfrac{1}{2}c^{iklr}e_{ik}e_{lr}, \tag{10.13}$$

with the corresponding stress-strain relations

$$\sqrt{(G)}\tau^{ik} = \sqrt{(g)}E^{iklr}\gamma_{lr}, \tag{10.14}$$

$$\sqrt{(G)}t^{ik} = c^{iklr}e_{lr}. \tag{10.15}$$

The strain energy in this case can be also put in the alternative form

$$\rho W = \tfrac{1}{2}\tau^{ik}\gamma_{ik}. \tag{10.16}$$

§ 11. ELASTIC COEFFICIENTS.

In general, remembering the symmetric properties

$$c^{iklr} = c^{kilr} = c^{ikrl} = c^{lrik}, \tag{11.1}$$

we have 21 elastic coefficients c^{iklr} for an æolotropic body. If the coordinate system x_i is changed to another rectangular cartesian system z_i then the coefficients c^{iklr} become d^{iklr} and, in view of their tensor character,

$$d^{iklr} = \frac{\partial z_i}{\partial x_m}\frac{\partial z_k}{\partial x_n}\frac{\partial z_l}{\partial x_s}\frac{\partial z_r}{\partial x_t}c^{mnst}. \tag{11.2}$$

Symmetry of structure of the general æolotropic body will introduce relations among the elastic coefficients c^{iklr} which may be found directly by using (11.2) and suitable transformations of coordinates, but we will not examine these in detail here. When the body is homogeneous and isotropic the coefficients c^{iklr} are invariant under all transformations to other rectangular cartesian coordinates and it can be shown that the coefficients reduce to

$$c^{mnst}=\lambda\delta^{mn}\delta^{st}+\mu\delta^{ms}\delta^{nt}+\mu\delta^{mt}\delta^{ns}, \quad . \quad . \quad . \quad . \quad (11.3)$$

where λ, μ are constants. Hence †

$$\mathrm{E}^{iklr}=\lambda g^{ik}g^{lr}+\mu(g^{il}g^{kr}+g^{ir}g^{kl}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.4)$$

or

$$\mathrm{E}^{iklr}=\mu\left\{g^{il}g^{kr}+g^{ir}g^{kl}+\frac{2\eta}{1-2\eta}g^{ik}g^{lr}\right\}. \quad . \quad . \quad . \quad (11.5)$$

where

$$\lambda=\frac{2\eta\mu}{1-2\eta}:$$

From (10.14) and (11.5) we find that the stress-strain relation for an isotropic body takes the form

$$\sqrt{(\mathrm{G})}\tau^{ik}=\mu\sqrt{(g)}\left\{g^{il}g^{kr}+g^{ir}g^{kl}+\frac{2\eta}{1-2\eta}g^{ik}g^{lr}\right\}\gamma_{lr}, \quad . \quad . \quad (11.6)$$

or

$$\sqrt{(\mathrm{G})}\tau^{ik}=\mu\sqrt{(g)}\left\{g^{il}\gamma_l^k+g^{ir}\gamma_r^k+\frac{2\eta}{1-2\eta}g^{ik}\gamma_l^l\right\}. \quad . \quad . \quad (11.7)$$

If the coordinate curves θ_i are along the principal directions of strain then $\gamma_{ik}, g_{ik}, \mathrm{G}_{ik}, g^{ik}$, G^{ik} are all zero when $i \neq k$. Hence, when the stress–strain relations are of the form (11.6), $\tau^{ik}=0$ $(i \neq k)$, so that the principal directions of stress in the strained body B coincide with the principal directions of strain.

REFERENCES.

BIEZENO, C. B., and GRAMMEL, R., 1939, *Technische Dynamik*. Springer, Berlin.

BRILLOUIN, L., 1938, *Les Tenseurs en Mécanique et en Élasticité*, Masson, Paris, p. 226.

COURANT, R., and HILBERT, D., 1931, *Methoden der Mathematischen Physik*. Vol. 1. Springer, Berlin.

DEUKER, E. A., 1941, *Deutsche Mathematik*, **5,** 546.

GLEYZAL, A., 1949, *Quart. App. Math.*, **6,** 429.

LOVE, A. E. H., 1927, *A Treatise on the Mathematical Theory of Elasticity*. Cambridge University Press.

MILNE-THOMSON, L. M., 1948, *Proc. 7th International Congress for Applied Mechanics*, **33,**

MURNAGHAN, F. D., 1937, *Amer. Jour. Math.*, **59,** 233.

NEUBER, H., 1943, *Z. angew. Math. Mech.*, **23,** 321.

RICHTER, H., 1948, *Z. angew. Math. Mech.*, **28,** 205.

RICHTER, H., 1949, *Z. angew. Math. Mech.*, **29,** 65.

TREFFTZ, E., 1928, *Handbuch der Physik*, Vol. 6. Springer, Berlin 1939.

---

† This result may also be obtained by assuming that the strain-energy function W is a function of the strain invariants. See, for example, F. D. Murnaghan (1937).

XXVII. *The Emission of Short-range Alpha Particles from Light Elements under Proton Bombardment.*—II. *Further Observations on the Reaction* $^{10}B(p\alpha)^{7}Be$.

By W. E. BURCHAM * and JOAN M. FREEMAN †.

[Received December 23, 1949.]

SUMMARY.

The energy release in the reaction $^{10}B(p\alpha)^{7}Be$ has been measured, using targets of known thickness and found to be $1{\cdot}147 \pm 0{\cdot}010$ keV. Evidence for the existence of an excited state of energy about 400 keV. in the ^{7}Be nucleus was obtained. The cross-section for the reaction was found by comparing the yield of alpha particles with the yield from the $^{10}B(n\alpha)^{7}Li$ reaction observed with the same target. A value of $4{\cdot}5 \times 10^{-27}$ cm.^{2} at 330 keV. proton energy was found.

§ 1. INTRODUCTION.

IN a previous paper (Burcham and Freeman 1949, subsequently referred to as I.) measurements of the energy release and cross-section for the reaction

$$^{10}B + {}^{1}H \rightarrow {}^{7}Be + {}^{4}He \quad . \; . \; . \; . \; . \; . \; . \; . \quad (1)$$

were reported. An improved cross-section determination has now been made by comparing the yield of reaction (1) with that of the slow neutron reaction

$$^{10}B + {}^{1}n \rightarrow {}^{7}Li + {}^{4}He. \quad . \; . \; . \; . \; . \; . \; . \; . \quad (2)$$

This work, together with further observations on the energy release and excitation function for reaction (1), is described in the present paper.

§ 2. ENERGY RELEASE IN THE REACTION $^{10}B(p\alpha)^{7}Be$.

(a) *Q-value for the ground-state transition.*

The energy release reported in I. was $1{\cdot}11 \pm 0{\cdot}06$ MeV., while in a similar investigation Chao, Lauritsen and Tollestrup (1949) found a value of $1{\cdot}148 \pm 0{\cdot}005$ MeV. The large error in the former determination arose partly from uncertainty in the thickness of the isotopic targets used, and in order to eliminate this error a special study of the alpha-particle group emitted from boron targets of different thickness has now been made. A set of targets each of area $0{\cdot}20$ cm.^{2}, and of thickness known from the ion current in the electromagnetic separator, was prepared; the density of the deposits ranged from 30 to 300 micrograms per cm.^{2}. These

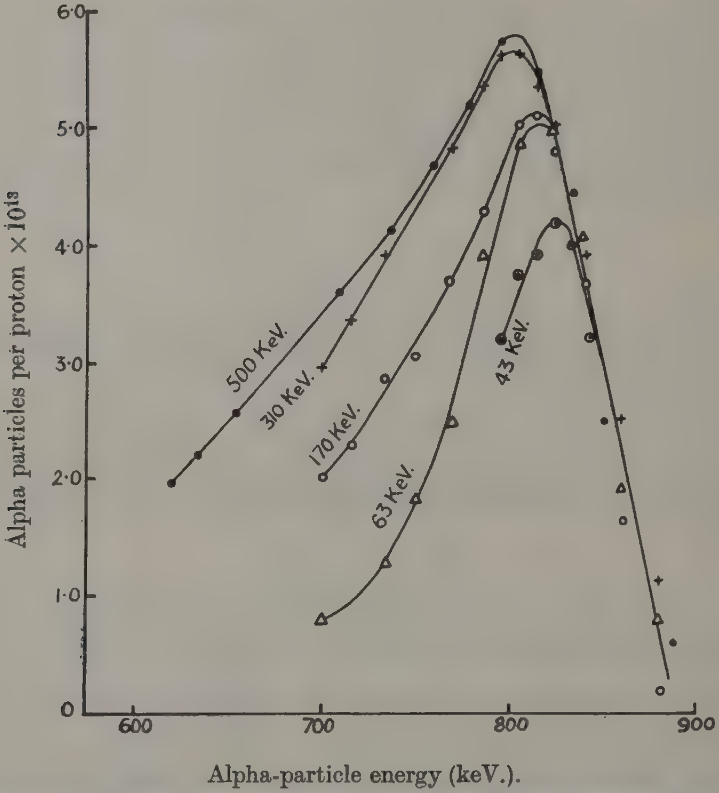
---

* Selwyn College, Cambridge; † Newnham College, Cambridge ‡.

‡ Communicated by the Authors.

targets were bombarded with 432 keV. protons, and the alpha particles emitted at 123° to the proton beam were observed in the magnetic analyser as described in I. Some of the curves thus obtained are shown in fig. 1; the energy scale for these curves was obtained from a calibration of the magnetic analyser carried out by means of protons of energy known to ±5 keV., scattered from a thin gold film as described in I. In the experiments with boron a fluorescent screen of circular aperture 1·0 cm. in diameter was used, and the resolving power of the analyser, as defined in I., was 25 (width 66 keV. for alpha particles of energy 830 keV.). From the known analyser resolution and the excitation function for the reaction a curve representing the energy spectrum to be expected from a target of given thickness was constructed. For this purpose it was assumed that the alpha particles emerging from a particular depth in the target gave a spectrum which could be approximated by a

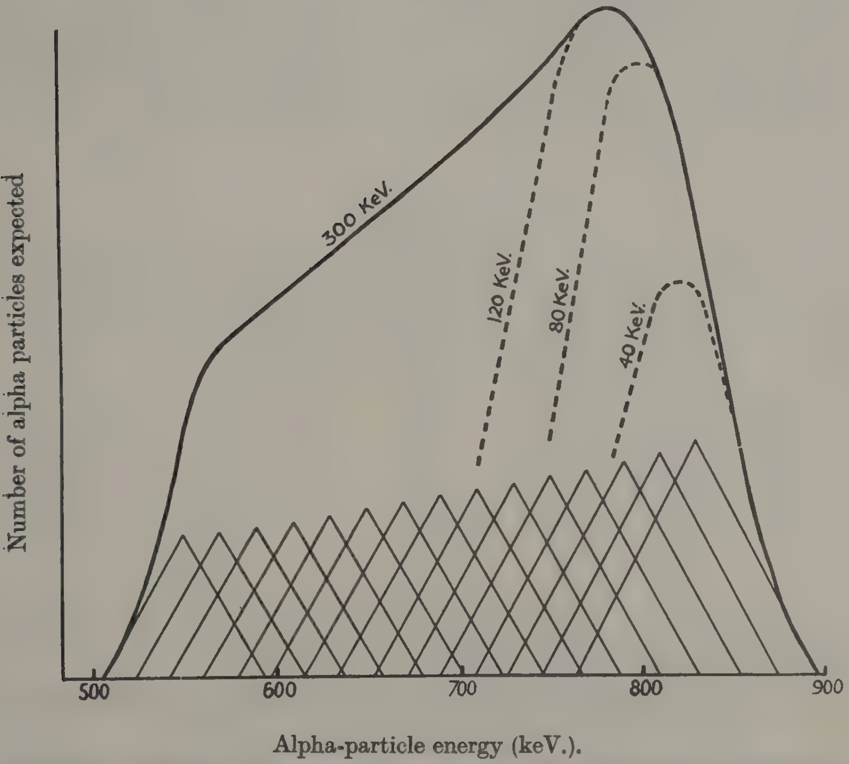
Fig. 1.



Alpha-particles from the reaction $^{10}B(p\alpha)^7Be$. Proton energy 432 keV.; angle of observation 123°; target thicknesses 43, 63, 170, 310 and 500 keV. for alpha particles.

triangle of height proportional to the cross-section for the reaction at the average proton energy for this depth in the target, and of width (at half-maximum) determined by the resolution of the analyser. A series of such triangles was drawn with 20 keV. intervals, the first representing alpha particles coming from an average depth of 10 keV., the second representing alpha particles from a depth of 30 keV., and so on, for a target of total thickness equivalent to 300 keV. (0·2 mgm./cm.^{2}), as shown in fig. 2. An integrated curve which would be expected to represent the observed spectrum was thus obtained (full curve). Similar curves, shown dotted in fig. 2, were drawn for targets of equivalent thickness 40, 80 and 120 keV. The intensity falls off on the left-hand side of the peak due to the effects of the exponential form of the excitation function and of the decreasing resolution width. It can be seen that for targets of equivalent thickness 120 keV. or more the peak of the spectrum occurs at an energy 60 keV. less than that for an infinitely thin target, and that only the first 120 keV. equivalent thickness of the target contributes to the peak.

Fig. 2.



Reaction $^{10}B(p\alpha)^7Be$. Theoretical alpha-particle distributions for targets of different thickness.

The position and intensity of the latter should thus be independent of the thickness for such targets. For thinner targets the intensity would be expected to decrease and the peak to move to the right. A comparison of the constructed curves and the experimental curves of fig. 1 shows that the latter set, although somewhat narrower, exhibit the general features expected, and justifies the procedure of deducing from the constructed curves the depth in the target in keV. equivalent for which the peak of the alpha-particle group from each of the experimental targets would be expected to occur. The stopping power of the boron targets for alpha particles of about 800 keV. energy was assumed to be 1450 keV. per mgm./cm.², which is consistent with the data of Madsen and Venkateswarlu (1948) on the loss of energy of protons in beryllium*. The results are set out in Table I., which gives the actual target thickness in mgm./cm.², the effective thickness allowing for inclination θ of the target to the proton beam, the corresponding thickness in keV. equivalent for alpha particles, the depth in the target for which the peak would be expected to occur, and the alpha-particle intensity and energy observed at the peak of each group.

TABLE I.

Target thickness			Depth of peak from surface (keV.) (theoretical)	Intensity at peak (observed) alpha-particles per 10^{13} protons	Energy at peak (observed) (keV.)
mgm./cm.²	$\frac{\text{mgm./cm.}^2}{\cos\theta}$	keV. equiv.			
0·300	0·345	500	60	5·7	805
0·185	0·215	310	60	5·6	804
0·165	0·190	270	60	4·9	814
0·105	0·120	170	60	5·1	815
0·038	0·044	63	32	5·0	819
0·026	0·030	43	22	4·2	825

When the alpha-particle energy is plotted against the theoretical depth of the peak from the surface (column 4) the result shown in fig. 3 is obtained. The extrapolation to zero target thickness gives an alpha-particle energy of 833 keV., with an uncertainty in the extrapolation of 3 keV. This figure must be further corrected to allow for the passage of the alpha particles through the thin layer of carbon deposited on the surface of the targets from pump-oil vapours. The thickness of this layer was found by determining the yield of radioactive N^{13}, formed in the $C^{12}(p\gamma)$ reaction, from the boron targets in comparison with that obtained from a thick carbon target. The average value of the thickness found for the boron targets was equivalent to 3±1 keV. for protons ; this correction has been made to the incident proton energy and a correction of 12±4 keV. to the emergent alpha-particle energy. With the resulting proton energy of 429±7 keV. and alpha-particle energy of 845±7 keV. at 123° the Q-value for the ground state transition in the $^{10}B(p\alpha)$ reaction is found to be

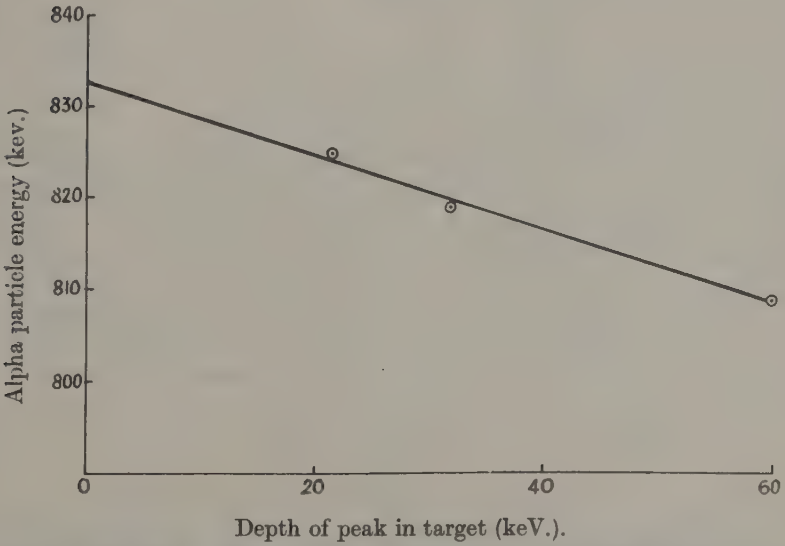
---

* *Note added in proof.*—The stopping power of beryllium for protons of velocity equal to that of 800 keV. alpha particles can be obtained directly from the recent work of Warsaw (1949, *Phys. Rev.*, **76,** 1759).

$$\underline{Q=1{\cdot}147\pm0{\cdot}010\text{ keV.},}$$

in good agreement with the work of Chao, Lauritsen and Tollestrup (1949). The error in this Q-value is less than that calculated directly from the stated errors in the alpha-particle and proton energies, since calibration errors affect these two energies in the same sense.

Fig. 3.



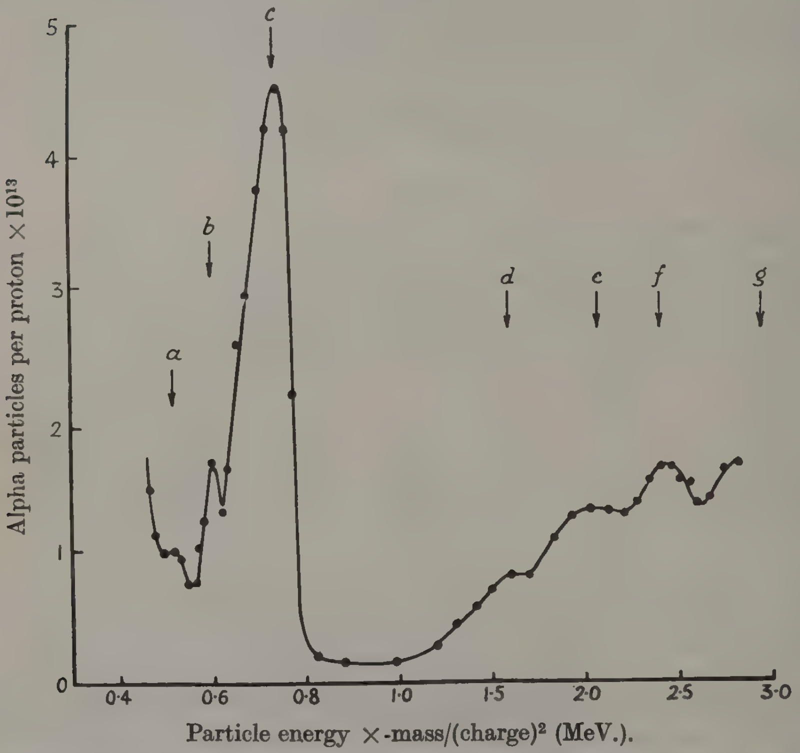
Reaction $^{10}B(p\alpha)^7Be$. Extrapolation of observed alpha particle energy to zero target thickness.

The curves of fig. 1 show that the shape of the high energy side of the alpha-particle distribution is independent of target thickness. An estimate of the Q-value from an " extrapolated " energy derived by drawing a tangent to the observed spectrum gave 1·157±0·020 MeV. ; this is less accurate than the preceding value owing to greater uncertainty in the analyser calibration for " extrapolated " ranges.

(b) *The first excited state of* ^7Be.

It was pointed out in I. that the similarity between the nuclei ^7Li and ^7Be suggested that there might be a low-lying state of ^7Be analogous to the 480 keV. state of ^7Li. A search (I. fig. 10) for alpha particles emitted in the $^{10}B(p\alpha)^7Be$ reaction which would correspond to transitions to such an excited state showed that the intensity of such a group was less than 10 per cent of that of the main group for a proton energy of 230 keV. If, however, the *a priori* probability of emission of the two groups is the same, barrier penetrability would reduce the intensity of an alpha-particle group corresponding to an excited state of more than 400 MeV. excitation to about this fraction of the intensity of main group. The use of a much higher bombarding voltage to increase the yield was impossible since the alpha particles would have been obscured by protons scattered from the target backing of thick copper, and targets of the ^{10}B isotope deposited on thin foils were not available. In order to extend the work to higher voltages a search was made for singly charged alpha particles corresponding to the excited state transition ; fig. 4 shows the results obtained with a target of thickness 0·165 mgm./cm.^{2}, a proton energy of 430 keV., an angle of observation of 135°, and an analyser resolving power of 50. The analyser was calibrated at the higher energy values by observations of the protons from the $^{12}C(dp)^{13}C$ reaction (Q=2·70 MeV.). The distribution

Fig. 4.



Reaction $^{10}B(p\alpha)^7Be$. Singly and doubly charged groups observed at 135° with E_p=430 keV. and resolution of 4 per cent in energy. Note change of scale at 1·0 MeV.

of particles observed with high analyser fields could not be completely resolved into separate groups, owing probably to the retardation of the recoil groups in the target. If, however, correction is made, as described in § 2, for the effective depth of the target and for the presence of a carbon layer, the expected positions of the singly and doubly charged alpha particles and singly and doubly charged ^{7}Be nuclei corresponding to the ground state transition are as marked (g, c, f, b). The groups (e, d) could be due to singly charged alpha particles and Be nuclei corresponding to a transition to an excited state of about 400 keV., and there is some slight evidence for an associated ^{4}He^{++} group (a). In confirmation of this interpretation it was shown that all the groups except c and g (of energy 740 keV.) disappeared when the zinc sulphide screen was covered by a 1·5 mm. foil, which indicated that groups a, b, d, e, f were of lower energy. Groups f and e were found to move with increasing bombarding proton energy in the way expected of ^{7}Be$^+$ and ^{4}He$^+$ nuclei respectively.

At this time it was learned that definite evidence for the existence of a 435 keV. state in ^{7}Be had been obtained by Brown, Chao, Fowler and Lauritsen* with better experimental conditions, and it was felt that nothing further could be gained by continuing work on the low energy group with the available targets. The evidence for the existence of an excited state of ^{7}Be near 400 keV. is in agreement with the results of Brown *et al.*; the ratio of intensities of the transitions to the ground and excited states for a proton energy of 430 keV. cannot be estimated accurately from the present results, but appears to be about 5 to 1.

The existence of a 435 keV. state in ^{7}Be should lead to the emission of two groups of neutrons in the reactions

$$^7\mathrm{Li}(pn)^7\mathrm{Be} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (3)$$

and

$$^6\mathrm{Li}(dn)^7\mathrm{Be} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (4)$$

and to gamma radiation of energy 430 keV. in these two reactions and in the ^{10}B$(p\alpha)^7$Be reaction. Gamma radiation of this energy has been observed in the $(p\alpha)$ reaction and in reaction (4) by Lauritsen and Thomas*, and in the $(p\alpha)$ reaction in this laboratory by Jelley and Freeman (unpublished). The observations of Freier, Lampi and Williams (1949) and of Taschek and Hemmendinger (1948) revealed no second neutron group in reaction (3) in intensity exceeding 10 per cent of the main group. Recently, however, evidence for neutron groups in this reaction corresponding to levels in ^{7}Be of energies $0{\cdot}206 \pm 0{\cdot}07$, $0{\cdot}470 \pm 0{\cdot}07$ and $0{\cdot}745 \pm 0{\cdot}07$ has been obtained by Grosskreutz and Mather (1949). The 745 keV. state would not have been observed in the present $(p\alpha)$ experiments, and the 470 keV. state seems to correspond to that found by Brown *et al.* and also indicated in the present work. Alpha particles corresponding to a transition to a 205 keV. state would have been observed in approximately the position occupied by group *b*, which was interpreted as $^7Be^{++}$ recoils in fig. 4. The present work does not definitely establish the identity of group *b*, but group *f*, which was interpreted as the corresponding $^7Be^+$ group, was shown to increase in energy with increasing bombarding energy as expected for particles of this mass and charge.

---

* We are indebted to Professor T. Lauritsen for sending us this information in advance of publication.

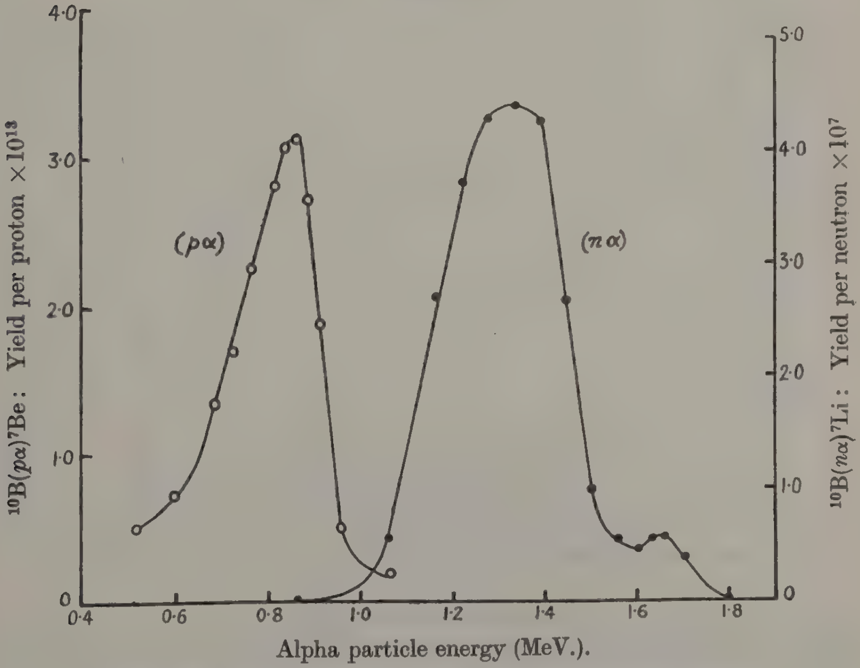
Evidence for a second neutron group in reaction (4) has been obtained in the Cavendish Laboratory by Gibson (in course of publication) using the photographic plate technique. The neutron spectrum is consistent with an energy level of 435 keV. No evidence for a 745 keV. level was obtained by Mandeville, Swan and Snowdon (1949) in their study of this reaction.

§ 3. COMPARISON OF THE REACTIONS $^{10}B(p\alpha)^7Be$, $^{10}B(n\alpha)^7Li$.

(a) *Energy release.*

For the observation of reaction (2) the magnetic analyser and target chamber were transported to the Atomic Energy Research Establishment, Harwell, and set up in such a way that a beam of thermal neutrons could pass through the ^{10}B target. The region of the target of area 0·31 cm.^{2} which had previously been bombarded by protons was defined by a slight discoloration and a screen was placed over the target for the $(n\alpha)$ experiments, so that only the alpha particles emitted from the area thus defined could be observed. A foil of stopping power 4 mm. was placed over the zinc sulphide screen to prevent the detection of the $^7Li^{++}$ nuclei produced in the reaction, while not appreciably reducing the efficiency for alpha-particle detection. The analyser resolving power was 25 and an angle of observation of 83° was used. The energy spectrum obtained is shown in fig. 5, curve $(n\alpha)$. The two alpha-particle groups, corresponding to transitions to the ground state and to the first excited state, are clearly evident. The main group in the $(n\alpha)$ experiment is appreciably broader than that found in $(p\alpha)$ experiments, and this can be explained on the basis of target thickness effect by reference to a curve constructed by a procedure similar to that described for the derivation of fig. 2. In the $(n\alpha)$ case the cross-section for the reaction is the same at all depths in the target, so that the triangles representing the contribution to the total spectrum of the alpha-particles arising from various depths all have the same height, decreasing in width only with decreasing energy. The curve constructed on this basis is shown in fig. 6. It is seen that contributions to the peak are received from alpha particles emitted in the first 200 keV. equivalent thickness of the target, and the peak occurs at an energy 110 keV. below that of alpha particles emitted from the surface. A comparison of this curve with that of fig. 5 shows that, in the position used, the target was effectively about 200 keV. " thick ", and that a correction of 110 keV. should therefore be applied to the energy at the peak of the spectrum. The energy of the main alpha-particle group emitted from the surface of the target was thus deduced as 1·45±0·02 MeV. The effective thickness of the carbon film on the target was estimated, as previously described, to be 20 keV. for alpha particles, giving a true particle energy of 1·47±0·02 MeV. This gave the energy release in the ^{10}B($n\alpha$)^{7}Li* reaction as Q=2·31±0·04 MeV. From the position of the alpha-particle group due to the ground state transition the true energy of these particles was estimated to be 1·78±0·03 MeV., giving an energy release of 2·80±0·05 MeV. These values are in agreement with previous work (Boggild 1945, Gilbert 1948, Chao, Lauritsen and Tollestrup 1949). The accuracy of the present determination was limited by the intensity of the slow neutron beams available and by the solid angle of the analyser.

Fig. 5.



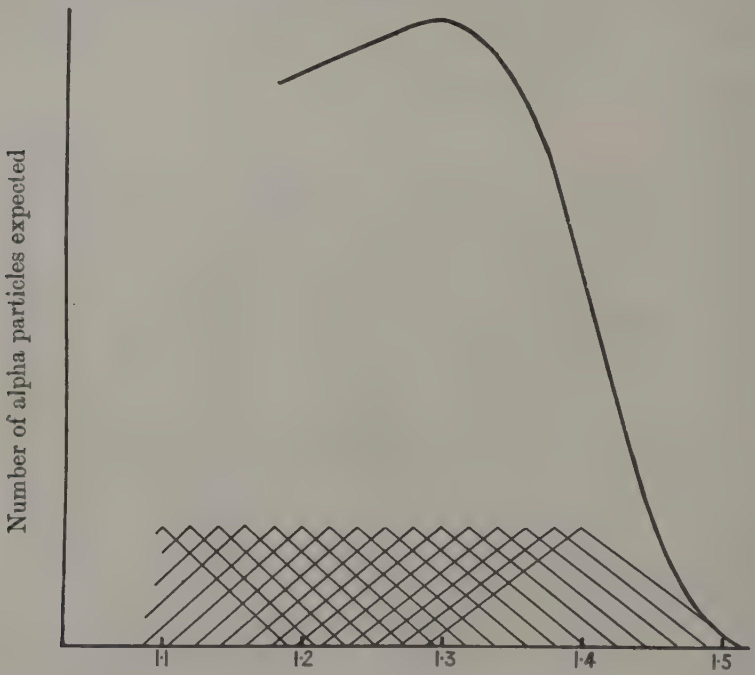
Reactions ^{10}B($p\alpha$)^{7}Be and ^{10}B($n\alpha$)^{7}Li. Alpha-particle groups.

The ^{10}B target used in this ($n\alpha$) experiment was bombarded with protons with the main object of comparing the yields from the two reactions, but the results also allowed another check on the energy release. A proton energy of 330 keV. was used, and an angle of observation of 83°. The energy spectrum (fig. 5, curve ($p\alpha$)) indicated that the effective target thickness was greater than 120 keV., as expected from the ($n\alpha$) result, so that a correction of 60 keV. for target thickness was necessary. An allowance of 20 keV. was made for the carbon film formed on the target, and the true alpha-particle energy was deduced as 935±16 keV. This gave a value of 1·14±0·03 MeV. for the energy release, in satisfactory agreement with the more accurate estimate given above.

(b) *Cross-sections.*

The ordinates of fig. 5 give directly the yield of alpha particles in the reactions (1) and (2) per incident proton or thermal neutron respectively. The yield for the proton reaction was obtained by integrating the charge collected on the target during the ($p\alpha$) experiments, care being taken to eliminate the effects of secondary electrons. The amplifier gain and bias setting were known (see I., § 2 (c)) to be suitable for the detection of alpha particles of 800 keV. with 100 per cent efficiency, and were high enough to make correction for the presence of the lower energy $^7Be^{++}$ nuclei unnecessary. In the ($n\alpha$) experiment the flux of neutrons incident on the target was measured by placing a calibrated manganese foil of area 1 cm.^{2} in the position of the target and allowing it to be irradiated for a known period. The beta-activity was found in a standard counting arrangement, whose efficiency was checked by means of a uranium source ; the neutron flux used was $6{\cdot}5\times10^6\pm5$ per cent. No correction

Fig. 6.



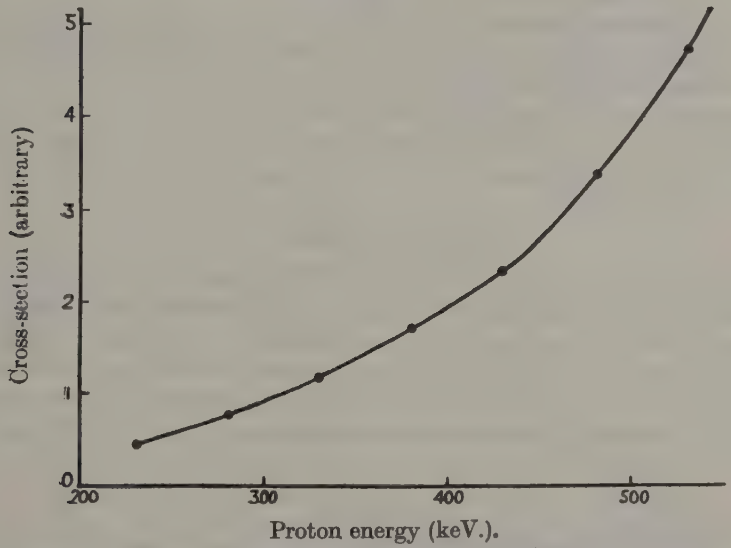
Alpha-particle energy (MeV.).

Reaction $^{10}B(n\alpha)^7Li$. Theoretical alpha-particle distribution for thick target.

for the presence of ^{7}Li recoils in reaction (2) was necessary, since an absorbing foil was used over the zinc sulphide screen in the ($n\alpha$) experiment.

The analyser resolution in both ($n\alpha$) and ($p\alpha$) experiments was less than the equivalent thickness of the target. The ratio of yields must therefore be corrected by a factor (1·57) equal to the ratio of the energies of the particle groups. A further correction for charge exchange of 1·27 for the ($n\alpha$) and 1·66 for the ($p\alpha$) reaction was deduced from the figures given by Briggs (1927). With these corrections, and after subtraction of background counts, the ratio of yields at 83° per incident particle becomes $1\cdot37\times10^{-6}$. This, assuming isotropic angular distributions, is taken to be the ratio of the cross-section for reaction (1) with 330 keV. protons to that of reaction (2) with thermal neutrons. The neutron flux measurement by means of the $1/v$-law manganese detector assumed a neutron velocity of $2\cdot5\times10^5$ cm. per sec., and the ^{10}B cross-section for neutrons of this velocity is 3285 barns, from the results of Fermi, Marshall and Marshall (1947). The ($p\alpha$) cross-section is, therefore,

Fig. 7.



Reaction ^{10}B($n\alpha$)^{7}Be excitation function. Absolute cross-section at 330 keV.$=4\cdot5\times10^{-27}$ cm.^{2}

$$\sigma_{p\alpha}\,(330\text{ keV.})=4\cdot5\times10^{-27}\text{ cm.}^2$$

In order to compare this figure with that previously given for a proton energy of 530 keV. in I., an extension of the excitation function given in I., fig. 11, was made. This is given in fig. 7, the ordinate scale being fixed by the 330 keV. determination, and from this

$$\sigma_{p\alpha}(530\text{ keV.})=1\cdot9\times10^{-26}\text{ cm.}^2,$$

which is to be compared with the value of $1 \cdot 2 \times 10^{-26}$ cm.^{2} given in I. The difference is not surprising in view of the uncertainty of the factors which entered into the earlier value ; in the present method knowledge of the solid angle of the analyser is not required, and the exact target thickness need not be known since the analyser selects a definite layer of the target determined by the resolution. The calculation of the cross section given in I. from the yields shown as ordinates in I., fig. 11, included a large correction for the ratio of areas of the proton beam and the boron target; this correction is unnecessary in the present experiment. The main source of experimental error, which has not been eliminated in the present cross-section determination, is a possible over-estimate of the counting efficiency for alpha particles of 3·5 mm. range emerging from the stopping foil in the $(n\alpha)$ reaction. From the curves given in I., fig. 4 (*b*) (bias 20), this error is unlikely to exceed 20 per cent.

As a further check the cross-section for the $(p\alpha)$ reaction was also deduced from the observations on small weighed targets described in § 2. The solid angle of the analyser and the target depth determined by resolution were assumed and a value of $2 \cdot 0 \times 10^{-26}$ cm.^{2} at $E_p = 530$ keV. was found, in agreement with that deduced from the $(n\alpha)$ reaction within the limits of error of the measurements.

ACKNOWLEDGMENTS.

We wish to thank Sir John Cockcroft for granting us permission to use the facilities at the Atomic Energy Research Establishment, Harwell, and Mr. F. W. Fenning for much assistance there. We are grateful to Dr. W. D. Allen, of the Atomic Energy Research Establishment, Harwell, for the boron isotopes. We also wish to thank Messrs. W. Birtwhistle and D. D. Stewart for much technical assistance.

REFERENCES.

BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 69.
BOGGILD, J. K., 1945, *Kgl. Dansk. Vidensk. Selsk. Mat.-fys. Medd.*, **23**, No. 4.
BRIGGS, G. H., 1927, *Proc. Roy. Soc.* A, **114**, 341.
BURCHAM, W. E., and FREEMAN, J. M., 1949, *Phil. Mag.*, **40**, 807.
CHAO, C. Y., LAURITSEN, C. G., and TOLLESTRUP, A. V., 1949, *Phys. Rev.*, **76**, 586.
FERMI, E., MARSHALL, J., and MARSHALL, L., 1947, *Phys. Rev.*, **72**, 193, 703 *b*.
FREIER, G., LAMPI, E. E., and WILLIAMS, J. H., 1949, *Phys. Rev.*, **75**, 901.
GILBERT, C. W., 1948, *Proc. Camb. Phil. Soc.*, **44**, 447.
GROSSKREUTZ, J. C., and MATHER, K. B., 1949, *Bull, Am. Phys. Soc.*, **24**, No. 7, 15.
MADSEN, C. B., and VENKATESWARLU, P., 1948, *Phys. Rev.*, **74**, 648.
MANDEVILLE, C. E., SWAN, C. P., and SNOWDON, S. C., 1949, *Phys. Rev.*, **76**, 980.
TASCHEK, R., and HEMMENDINGER, A., 1948, *Phys. Rev.*, **74**, 373.

XXVIII. *Relaxation Methods applied to Problems of Plastic Flow* *.—
I. *Notched Bar under Tension.*

By J. A. JACOBS,
Royal Holloway College, University of London †.

[Received January 3, 1950.]

SUMMARY.

Orthodox analysis of most problems of plastic flow is extremely difficult, in that the governing equations are not linear, so that partial solutions may not be superposed. Many problems, however, yield to treatment by relaxation methods. This paper considers the consequence of a particular hypothesis regarding plastic strain, viz. the Mises-Hencky criterion, and, by application of a numerical method due to Allen and Southwell, traces the effect of plastic straining. The success of the method in no way depends upon the particular yield hypothesis adopted. Three problems are considered:

(i) A two-dimensional slit tension member.
(ii) A sharp-ended bar under specified edge tractions.
(iii) The uniform compression of a block between two smooth, rigid plates which do not overlap it.

In each case, results and conclusions are summarized and presented graphically.

§ 1. INTRODUCTION.

ORTHODOX analysis of most problems of plastic flow is extremely difficult in that the governing equations are not linear, so that partial solutions may not be superposed. Moreover, in many cases the shape of the boundary and the boundary conditions are such as to preclude an exact solution, even in the elastic case. Many of these difficulties are resolved when the problem is attacked by relaxation methods—thus, for example, the shape of the boundary does not materially increase the complexity of the problem. Again, many of the exact solutions are extremely complicated, being usually expressed in series form. The convergence is often very slow, and the numerical value of the stresses at various points would require as much labour as a normal solution by relaxation methods.

Relaxation methods are peculiarly suited to a class of problem involving boundaries or interfaces not initially known—problems which, in many cases, present insuperable difficulties to an orthodox approach. In this connection, besides problems of plastic flow, relaxation methods have been successfully applied by Shaw and Southwell (1941) to problems relating to the percolation of fluids through porous materials, and by Southwell and Vaisey (1946) to fluid motions characterized by free streamlines.

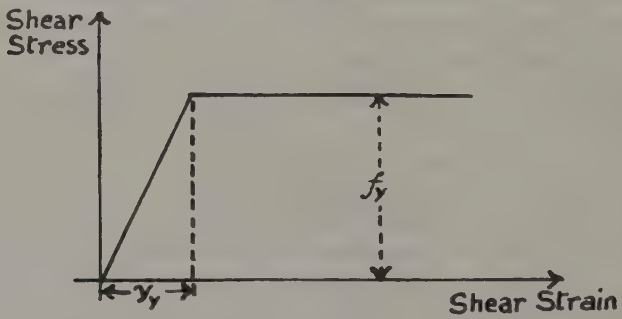
---

* This paper is based upon work forming part of the writer's Ph.D. thesis, London University, June 1949.

† Communicated by the Author.

The true criterion of elastic failure and the relation between stress and strain during plastic distortion in real materials has yet to be determined. Many hypotheses have been advanced, the validity of any of which only experiment can finally endorse. Direct experimental study of stress-strain relations is made more difficult by the fact that stress is not itself measurable. Experiment can only verify or deny results computed from one particular theory in regard to total displacements produced by imposed external tractions. We adopt a particular hypothesis regarding plastic failure and use it as a basis for our computations. It is to be emphasized that the methods evolved do not depend on the particular hypothesis adopted, but could be applied similarly with any other criterion.

Fig. 1.



§ 2. General Theory.

The first attack on plastic problems by relaxation methods was made by Christopherson and Southwell (1938), who considered the plastic torsion of a bar of equilateral triangular cross-section. Later Christopherson (1940) applied the same methods to I-section girders, and Shaw (1942) considered hollow (*i.e.* multiply-connected) sections, where plastic strain can commence at the external or at an internal boundary. All these problems of plastic torsion based their solution on a hypothesis put forward by Prandtl (1923) which assumes the relation between shear stress and shear strain to be that indicated in fig. 1.

For stresses below some limiting value f_y the material behaves elastically, but the stress cannot exceed this limiting value f_y with which any shear strain in excess of a limiting elastic shear strain γ_y can be associated. This imposes (in the torsion problem) an upper limit upon the gradient of the stress function. This restriction can be represented by what is usually known as the Prandtl roof, and the ensuing computation is straightforward.

The next major advance was made by Allen and Southwell (1950), who investigated more complex problems of plastic straining in two-dimensional stress systems. In particular, they considered notched tensile specimens having a central waist. Two examples were taken, in which the notches were semicircular and sharp re-entrant right angles. One of the examples we have considered is a limiting case of this type of problem—a slit tension member.

For the more complex stress systems considered, a generalization of the Prandtl hypothesis is required. The generalization we adopt rests on two assumptions :

(i) The first occurrence of plastic straining is determined by the "Mises-Hencky" criterion, *i.e.* the normal elastic relations hold until the principal stresses p_1, p_2, p_3 attain such values that

$$(p_2-p_3)^2+(p_3-p_1)^2+(p_1-p_2)^2=6k^2, \quad . \quad . \quad . \quad . \quad (1)$$

k being a constant for any material. The relation (1) holds thereafter throughout any region of plastic strain.

(ii) Where the material has yielded plastically, the strain consists of both an elastic and a plastic component. For the elastic part the usual stress-strain relations hold, viz.,

$$\mathbf{e}_1=\frac{1}{\mathrm{E}}\{p_1-\sigma(p_2+p_3)\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

etc., where E (Young's modulus) and σ (Poisson's ratio) have normal values.

The plastic part is assumed to be equivoluminal, *i. e.*

$$\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and in it the incremental shear strains resulting from the three principal shear stresses are directly proportional to these stresses and occur on the same three planes, *i. e.* if q_1, q_2, q_3 denote the prinicpal shear stresses, and $\Delta\gamma_1$,$\Delta\gamma_2$, $\Delta\gamma_3$ the principal incremental shear strains, then

$$\Delta\gamma_1 : \Delta\gamma_2 : \Delta\gamma_3=q_1 : q_2 : q_3. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Then if $\Delta\mathbf{e}_1$, $\Delta\mathbf{e}_2$, $\Delta\mathbf{e}_3$ are the principal incremental plastic strains, (4) may be written

$$(\Delta\mathbf{e}_2-\Delta\mathbf{e}_3):(\Delta\mathbf{e}_3-\Delta\mathbf{e}_1):(\Delta\mathbf{e}_1-\Delta\mathbf{e}_2)=(p_2-p_3):(p_3-p_1):(p_1-p_2). \quad (5)$$

where $\Delta\mathbf{e}_1$, $\Delta\mathbf{e}_2$, $\Delta\mathbf{e}_3$ have the same directions as p_1, p_2, p_3. Equations (4) and (5) may be written

$$\frac{\Delta\gamma_1}{q_1}=\frac{\Delta\gamma_2}{q_2}=\frac{\Delta\gamma_3}{q_3}=3\varDelta\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and

$$\frac{(\Delta\mathbf{e}_2-\Delta\mathbf{e}_3)}{p_2-p_3}=\frac{(\Delta\mathbf{e}_3-\Delta\mathbf{e}_1)}{p_3-p_1}=\frac{(\Delta\mathbf{e}_1-\Delta\mathbf{e}_2)}{p_1-p_2}=\tfrac{3}{2}\varDelta\lambda, \quad . \quad . \quad (7)$$

where $\Delta\lambda$ is an incremental quantity.

Thus $$2\Delta\mathbf{e}_1-\Delta\mathbf{e}_2-\Delta\mathbf{e}_3=\tfrac{3}{2}\Delta\lambda(2p_1-p_2-p_3),$$

i. e., using (3),

$$\Delta\mathbf{e}_1=\Delta\lambda\{p_1-\tfrac{1}{2}(p_2+p_3)\}, \qquad (8)$$

Thus, as regards the plastic component of the distortion, the stress-*incremental* strain relations have normal forms, the effective values of Young's modulus and Poisson's ratio being $\frac{1}{\Delta\lambda}$ and $\frac{1}{2}$ respectively.

The total strains associated with given stresses are given by

$$e_1=\mathbf{e}_1+\mathbf{e}_1$$

etc. That an exact analysis is rendered difficult is shown from the fact that (2) relates to *total* and (8) to *incremental* strains, both expressible in terms of total stresses.

Combining (2) and (8) gives three equations of the type

$$\Delta e_1=\frac{1}{\mathrm{E}}\{\Delta p_1-\sigma(\Delta p_2+p_3)\}+\Delta\lambda\{p_1-\tfrac{1}{2}(p_2+p_3)\}. \qquad (9)$$

Allen and Southwell (1950) deduce from the Mises-Hencky criterion (1) that for plane stress

$$\mathrm{X}_x^2-\mathrm{X}_x\mathrm{Y}_y+\mathrm{Y}_y^2+3\mathrm{X}_y^2\leqslant 3k^2, \qquad (10)$$

whilst for plane strain

$$(\mathrm{X}_x-\mathrm{Y}_y)^2+4\mathrm{X}_y^2\leqslant 4k^2, \qquad (11)$$

the sign of inequality relating to elastic conditions and the sign of equality to conditions of plastic straining. The condition (10) follows immediately, but (11) is obtained on the simplifying assumption that $\sigma=\frac{1}{2}$, whatever value may be appropriate in (9). The error due to this assumption is small.

§ 3. The Computation of the Stresses.

In plane strain (and in plane stress), provided that there are no body forces, all of the conditions of stress-equilibrium may be satisfied by writing

$$\mathrm{X}_x=\frac{\partial^2\chi}{\partial y^2},\quad \mathrm{X}_y=\frac{-\partial^2\chi}{\partial x\,\partial y},\quad \mathrm{Y}_y=\frac{\partial^2\chi}{\partial x^2} \qquad (12)$$

whatever be the nature of the accompanying strains. When known tractions act at a single boundary, χ and $\frac{\partial\chi}{\partial\nu}$ have the same boundary values whether plastic straining has occurred or not.

For specified edge tractions N_ν, N_s we have that

$$\begin{aligned}N_\nu \cos(x,\nu)-N_s \sin(x,\nu)&=X_\nu\\ &=X_x\cos(x,\nu)+X_y\sin(x,\nu)\\ &=\left\{\cos(x,\nu)\frac{\partial}{\partial y}-\sin(x,\nu)\frac{\partial}{\partial x}\right\}\frac{\partial\chi}{\partial y}=\frac{\partial}{\partial s}\left(\frac{\partial\chi}{\partial y}\right)\end{aligned}$$

In a similar manner

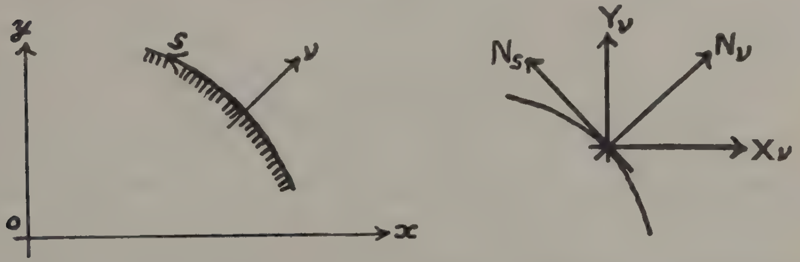
$$\begin{aligned}N_\nu \sin(x,\nu)+N_s \cos(x,\nu)&=Y_\nu\\ &=Y_y\sin(x,\nu)+X_y\cos(x,\nu)=-\frac{\partial}{\partial s}\left(\frac{\partial\chi}{\partial x}\right)\end{aligned}$$

Hence for all points of the boundary

$$-\frac{\partial\chi}{\partial x}=\int Y_\nu\,ds, \qquad \frac{\partial\chi}{\partial y}=\int X_\nu\,ds, \quad \dots\dots \quad (13)$$

both quantities being single-valued.

Fig. 2.



The specified boundary tractions can thus be reduced to given boundary values of $\frac{\partial\chi}{\partial x}$, $\frac{\partial\chi}{\partial y}$. Since the stress components depend on the derivatives of $\frac{\partial\chi}{\partial x}$, $\frac{\partial\chi}{\partial y}$, the lower limits of integration in (13) are arbitrary (so that $\frac{\partial\chi}{\partial x}$, $\frac{\partial\chi}{\partial y}$ may be given the value zero at some convenient point of the boundary).

Finally χ can be calculated at all points of the boundary from the equation

$$\chi=\int\left\{\frac{\partial\chi}{\partial x}\cdot\frac{dx}{ds}+\frac{\partial\chi}{\partial y}\cdot\frac{dy}{ds}\right\}ds, \quad \dots\dots \quad (14)$$

the lower limit again being arbitrary.

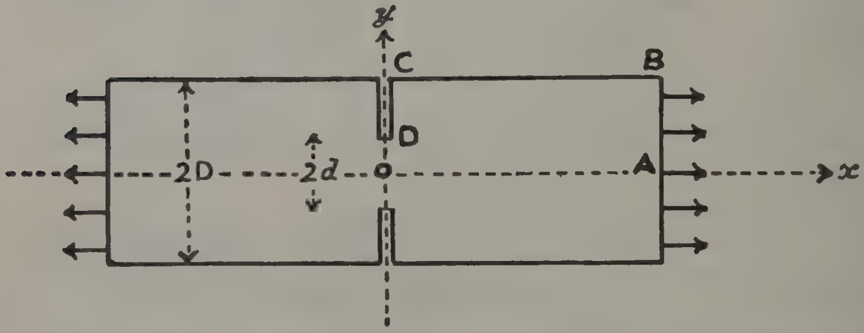
Writing (14) in the form

$$\begin{aligned}\chi&=x\frac{\partial\chi}{\partial x}+y\frac{\partial\chi}{\partial y}-\int\left\{x\frac{\partial^2\chi}{\partial s\,\partial x}+y\frac{\partial^2\chi}{\partial s\,\partial y}\right\}ds\\ &=x\frac{\partial\chi}{\partial x}+y\frac{\partial\chi}{\partial y}+\int(xY_\nu-yX_\nu)\,ds, \quad \dots\dots \quad (15)\end{aligned}$$

it follows that χ is single-valued at all points of the boundary.

Equations (13) and (15) enable boundary values of $\chi, \frac{\partial \chi}{\partial x}, \frac{\partial \chi}{\partial y}$ to be computed from specified edge tractions. So long as the whole of the plate remains elastic, the determination of χ at internal points is a biharmonic problem, whose method of solution by relaxation methods is given by Fox and Southwell (1945). Apart from a multiplying factor which increases linearly with the total load, the solution is unique. By substitution in the criterion (11) a value can be found for the greatest load that can be sustained elastically. If the load is increased still further, the criterion will be violated in some regions. In such regions the elastic solution must be modified to satisfy it, whilst leaving the values of $\chi, \frac{\partial \chi}{\partial x}, \frac{\partial \chi}{\partial y}$ unchanged on the boundary. This modification can be carried out by a relaxation method, described by Allen and Southwell (1950).

Fig. 3.



§ 4. PROBLEM (i.) STRESS DISTRIBUTION IN A TWO-DIMENSIONAL SLIT TENSION SPECIMEN.

Fig. 3 shows the slit tension specimen, subject to specified edge tensions. Three cases are considered, in which the ratio $\frac{D}{d}$, of the total width of the bar to the width at the waist is given the values 8, 4 and 2. From symmetry it will be sufficient to consider the quadrant OABCD. We can eliminate dimensional factors by writing

$$x = Dx', \qquad y = Dy', \qquad \chi = TD\chi', \quad . \quad . \quad . \quad . \quad . \quad (16)$$

2T denoting the total tension on the specimen.

Then such quantities as $\frac{\partial^2 \chi'}{\partial x'^2}$ measure stresses as multiples of $\frac{T}{D}$, the average tensile stress at the ends. Hereafter, however, the dashes attached to x, y and χ will be suppressed, and the symbols will stand for purely numerical quantities.

The boundary conditions are, along AB,

$$X_x = \frac{\partial^2 \chi}{\partial y^2}$$

is specified (and was taken as 3200 to obviate decimals).

$$X_y = \frac{-\partial^2 \chi}{\partial x\,\partial y} = 0 \tag{17}$$

along BC and CD, $X_\nu = Y_\nu = 0$, *i.e.*

$$\frac{\partial}{\partial s}\left(\frac{\partial \chi}{\partial x}\right) = \frac{\partial}{\partial s}\left(\frac{\partial x}{\partial y}\right) = 0 \quad . \; . \; . \; . \; . \; . \; . \tag{18}$$

so that $\frac{\partial \chi}{\partial x}$ and $\frac{\partial \chi}{\partial y}$ are both constant.

If we take $\frac{\partial \chi}{\partial x} = 0$ at D, then $\frac{\partial \chi}{\partial x} = 0$ along DC and CB, and hence $\frac{\partial \chi}{\partial x} = 0$ along AB (since $\frac{\partial \chi}{\partial x}$ is constant along BA and is thus equal to its value at B).

Since $\frac{\partial^2 \chi}{\partial y^2} = 3200$ along AB, $\frac{\partial \chi}{\partial y} = 3200y$, taking $\frac{\partial \chi}{\partial y} = 0$ at A. Thus at B, $(y=1)$, $\frac{\partial \chi}{\partial y} = 3200$, and hence $\frac{\partial \chi}{\partial y} = 3200$ along BC and CD.

Further, taking $\chi = 0$ at A,

$$\chi = 1600y^2 \quad . \; . \; . \; . \; . \; . \; . \; . \; . \tag{19}$$

along AB. Finally, from (14),

$$\chi = \int \left\{ \frac{\partial \chi}{\partial x} \cdot \frac{dx}{ds} + \frac{\partial \chi}{\partial y} \cdot \frac{dy}{ds} \right\} ds$$

$$= \int \frac{\partial \chi}{\partial y} \cdot \frac{dy}{ds}\, ds$$

along BCD where $\frac{\partial \chi}{\partial x} = 0$

$$= \int 3200\, dy$$

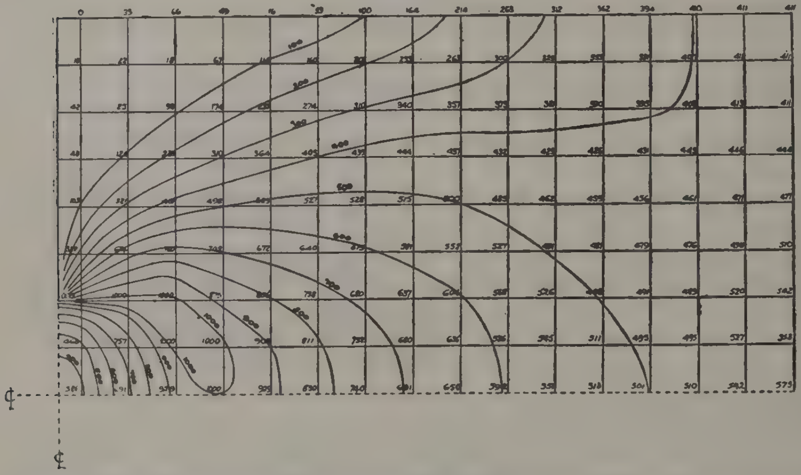
$$= 3200y - 1600 \quad . \; . \; . \; . \; . \; . \; . \; . \; . \tag{20}$$

since $\chi = 1600$ at B $(y=1)$. Actually along BC $(y=1)$, $\chi = 1600$, and along CD, $\chi = 3200y - 1600$.

The case $\frac{D}{d} = 4$ was first attacked. After a preliminary investigation on a coarse mesh, a finer net $(h = \frac{1}{8})$ was used.

The computation of χ and the stresses in the elastic case was carried out by the relaxational methods given by Fox and Southwell (1945). From this elastic solution we have to determine what modifications are necessary when, in a region of plastic straining, the condition (11) supervenes. It appears that a plastic enclave * will start at the bottom of the slit and extend into the specimen ; within an enclave, the initial (elastic) distribution of χ has to be modified in accordance with (11). Allen and Southwell (1950) give a relaxational method for the computation of χ within a plastic enclave. The process is rather laborious in that the relaxation pattern differs from point to point, and strictly only applies to infinitesimal increments in χ. Much labour can be saved by the use of inexact patterns, *exact* account, of course, being kept of the residuals.

Fig. 4.



Values of the Mises-Hencky equivalent stress (taking 1000 as its maximum value) when the plastic enclave has spread right across the bar. Case $\frac{D}{d}=4$.

In all the examples considered here, a modified relaxation technique was adopted—inexact patterns being used to give the general trend of the χ distribution, the final values being obtained by a systematic trial-and-error method. The labour involved was not prohibitive. The determination of the growth of the plastic enclave is facilitated by the computational device adopted of proceeding from step to step by *lowering* the value of the maximum equivalent stress rather than by *increasing* the applied tension.

---

* Southwell has suggested the use of the word enclave to denote a region of plastic straining.

Four distinct stages in the growth of the plastic region were obtained. Fig. 4 records the values of the Mises-Hencky equivalent stress when the plastic enclave has spread until it extends right across the bar. Fig. 5 summarizes these four stages, and shows the growth of the enclave with increasing tension.

The same problem was considered for other values of $\frac{D}{d}$, the ratio of the total width of the bar to the width at the waist. The case $\frac{D}{d}=2$ exhibited the same general characteristics as the case $\frac{D}{d}=4$ already considered. The elastic solution was verified by photo-elastic methods, and excellent agreement was obtained in spite of the difficulty of obtaining a model with a " slit "—practical considerations necessitating the notch having a finite though small width. One difference in the case $\frac{D}{d}=2$ is the rapidity with which the plastic enclave spreads. Although the growth of the enclave exhibits the same characteristics as in the case $\frac{D}{d}=4$, only two stages in the growth were obtained, as compared with four in the previous case. Once plastic flow begins, the enclave spreads

Fig. 5.

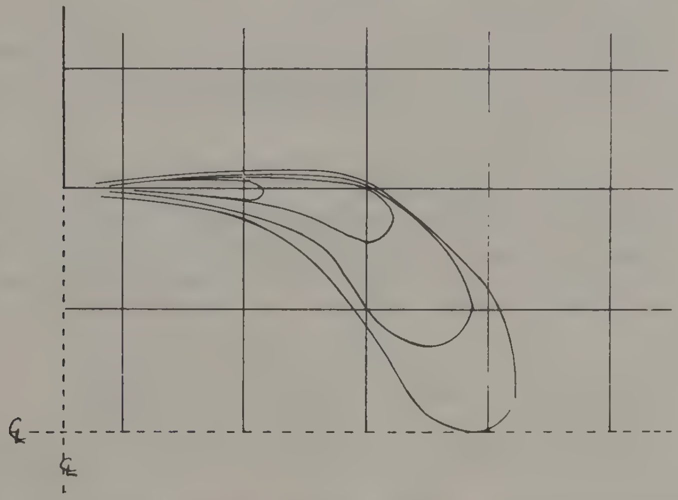
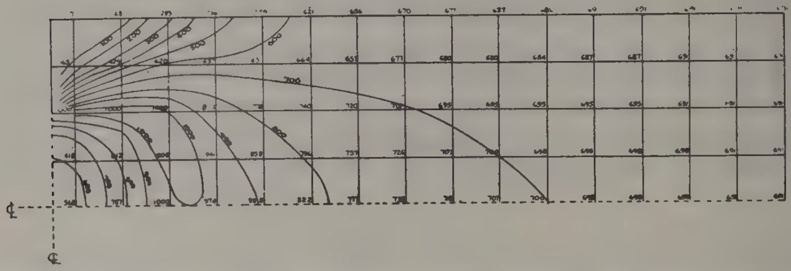


Diagram showing the progress of plastic straining—superposed plastic-elastic interfaces. Case $\frac{D}{d}=4$.

extremely rapidly and extends right across the bar for a very small increase in tension. Fig. 6 gives the values of the Mises-Hencky equivalent stress when the plastic enclave has spread right across the bar.

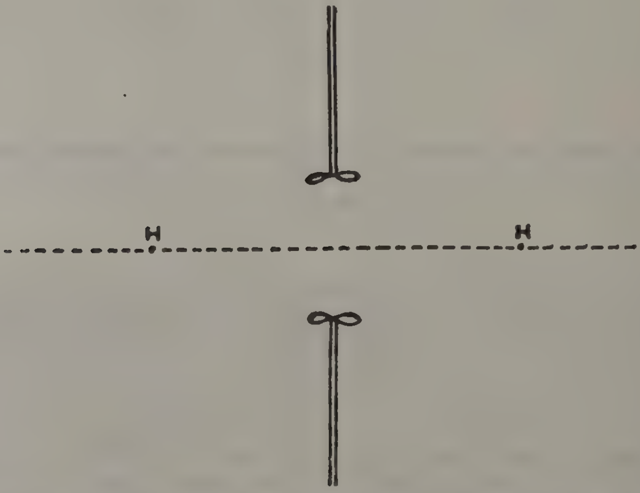
Fig. 6.



Values of the Mises-Hencky equivalent stress (taking 1000 as its maximum value) when the plastic enclave has spread right across the bar. Case $\frac{D}{d}=2$.

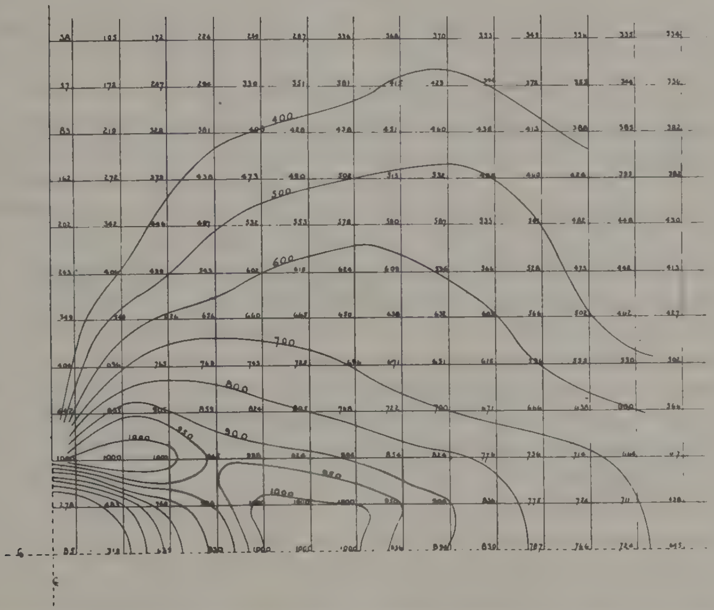
The case $\frac{D}{d}=8$ was an attempt to solve the problem for an infinite plate. There was a marked distinction in this case as compared with the other two cases. As plastic flow begins, an enclave, of the same shape as before, forms at the root of each notch and spreads into the plate ; this

Fig. 7.



is accompanied a little later by two new enclaves starting from points H, H on the centre line near the waist (see fig. 7). This phenomenon was also found by Allen and Southwell (1950) in the case of notches which were semicircular. The enclaves extend extremely rapidly and, as in the case $\frac{D}{d}=2$, only one intermediate stage was obtained before the two enclaves coalesced. Fig. 8 gives the value of the Mises-Hencky equivalent stress for the most advanced stage of plastic straining obtained. As the ratio $\frac{D}{d}$ increases, so the labour of computation greatly increases, because the number of mesh points rapidly multiplies. The case $\frac{D}{d}=8$ should give a good indication of the results for an infinite plane, as there was little variation in the stress distribution a long way from the slit. Preliminary investigation for the case $\frac{D}{d}=10$ showed but slight variation from the case $\frac{D}{d}=8$.

Fig. 8.



Values of the Mises-Hencky equivalent stress (taking 1000 as its maximum value) for the final plastic stage. Case $\frac{D}{d}=8$.

§ 5. Summary of Results.

Hill (1949) considers the state of stress in the core of a notched bar under an applied load sufficient to cause plastic flow, the notch being of sufficient depth to ensure that the plastic region is localized in and around the core. If the curvature of the notch is greatest at the root, the yield limit is first attained at that point, and as the applied end load is further increased, Hill assumes that the plastic region will spread in some manner from the notch to the axis, and shows that this plastic region must be of a certain minimum extent when the bar first extends. The initial growth of the plastic region for the limiting case of a deep notch, viz. a slit, which has been considered here, is in a direction which could hardly have been forecast in advance, and is not towards the axis of the bar as suggested by Hill's diagrammatic figure. Nothing in the present paper, however, conflicts with his results and conclusions :—"So long as part of the core is not plastic, no displacement of the ends of the specimen is possible. Even when the plastic region has spread right across the core, it does not necessarily follow that deformation is possible, since the plastic material may be rigidly constrained by the adjacent bulk of non-plastic material". From a consideration of the equations of Saint Venant governing the flow velocities in the plastic region, Hill shows that no extension of the specimen is possible until the plastic region has spread sufficiently far to include the entire slip lines OS, OS′ from the centre O to the free surface of the notch. At the moment when extension first occurs, the plastic boundary must touch OS, OS′, or must pass through either S and S′, or through O. Hill's surmise that the plastic boundary actually passes through S and S′ is confirmed by the results of this work :—

"Many writers, including some of the most recent, such as Sokolowsky (1946), and van Iterson (1947), have suggested slip-line fields for various problems quite arbitrarily, with no adequate discussion of their validity. In some cases elementary considerations are sufficient to prove them wrong, but in others only detailed numerical calculations can decide. In many problems the slip-line field in the region of plastically deforming material may be surmised without a full solution for the stresses in both elastic and plastic regions". The work of Allen and Southwell (1950) and the results of the investigations presented here show that in many cases the actual plastic region develops in a direction which could hardly have been surmised in advance. Thus, although it may be easy to suggest apparently plausible slip-line fields for a number of problems, detailed numerical calculations constitute the only *certain* method of procedure. In a future paper the results will be given of the investigation of a sharp-ended bar under specified edge tractions and of the uniform compression of a block between two smooth, rigid plates which do not overlap it.

§ 6. ADDENDUM.

Since the above investigations were carried out, Hill has shown (1950) that in the plane deformation of an incompressible plastic-elastic material under a given loading-path the distribution of stress at any stage can only be found uniquely from the stress equations alone if the plastic-elastic boundary is not cut more than once by any slip-line (and if, further, unloading does not occur). His proof depends on the fact that the slip-lines in the plastic region are the characteristics of both the stress-rate equations and the velocity equations. It follows that the calculated plastic regions in this paper are, after a certain stage, possibly unreliable since theoretically an unlimited number could be found to satisfy the stress equations and boundary conditions. Uniqueness is obtained by choosing the plastic region so that a velocity solution can be found. The results given here, however, do exhibit correctly the growth of the plastic region in the initial stages, and it is not anticipated that the final results are much in error.

REFERENCES.

ALLEN, D. N. DE G., and SOUTHWELL, R. V., 1950, *Phil. Trans.* (to be published).
CHRISTOPHERSON, D. G., 1940, *Amer. J. Appl. Mech.*, **7,** A 1–4 (reprinted *Roy. Aero. Soc. J.*, **44,** 425–432).
CHRISTOPHERSON, D. G., and SOUTHWELL, R. V., 1938, *Proc. Roy. Soc.* A, **168,** 317–350.
FOX, L., and SOUTHWELL, R. V., 1945, *Phil. Trans.* A, **239,** 419–460.
HILL, R., 1949, *Quart. J. Mech. and Applied Math.*, **2,** 40–52.
HILL, R., 1950, *The Mathematical Theory of Plasticity* (to be published by the Clarendon Press, Oxford).
VAN ITERSON, F. K. TH., 1947, *Plasticity in Engineering* (Blackie).
PRANDTL, L., 1923, *Z. Angew. Math. Mech.*, **3,** No. 6, 442.
SHAW, F. S., 1942, *J. Inst. Eng. Australia*, **14,** 273–277.
SHAW, F. S., and SOUTHWELL, R. V., 1941, *Proc. Roy. Soc.* A, **178,** 1–17.
SOKOLOVSKY, W. W., 1946, *Theory of Plasticity* (Izd. Ak. Nauk., Moscow).
SOUTHWELL, R. V., and VAISEY, G., 1946, *Phil. Trans.* A, **240,** 117–161.

XXIX. *On Tolman's Equation describing the Thermal Equilibrium in a Gravitating Sphere of Fluid.*

By H. A. Buchdahl,
Physics Department, University of Tasmania *.

[Received November 9, 1949.]

Abstract.

It is shown that by adopting a form of the line element slightly different from that chosen by Tolman a first order equation for the dependence of proper temperature on gravitational potential in a gravitating fluid sphere appears in place of Tolman's second order equation. The problem suggested by him concerning the meaning of a certain constant of integration therefore does not arise.

§1. Introduction.

Tolman (1934) has investigated the distribution of temperature in a gravitating sphere of fluid by invoking the thermodynamic conditions for static equilibrium. Assuming the form of the line element to be

$$ds^2 = -\exp\{\mu(r)\}(dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2) + \exp\{\nu(r)\}\, dt^2, \quad (1.1)$$

he arrives at a differential equation of the *second* order, a first integral of which is

$$\frac{d \log T_0}{dr} = -\frac{1}{2}\frac{d\nu}{dr} - \frac{B \exp\{-\frac{1}{2}\mu - \frac{1}{2}\nu\} T_0}{r^2}, \quad (1.2)$$

where B is a constant of integration and T_0 is the absolute temperature as measured by a local observer at rest. In order to arrive at a solution of (1.2) in the form

$$T_0 \exp\{\tfrac{1}{2}\nu\} = T_0 \sqrt{(g_{44})} = \text{const.} \quad (1.3)$$

he sets B equal to zero " on physical grounds " and remarks that " it would of course be interesting also to investigate the possibility for solutions of physical interest in which the constant of integration was not taken to be zero ".

We shall show, however, that the appearance of the term involving B in (1.2) is spurious in the sense that by adopting a form of the line element slightly different from (1.1) there arises a *first* order equation which replaces Tolman's second order equation, and which has the integral (1.3). The question of having to assign a particular value to a constant of integration equivalent to B consequently never arises. It might be added that we are here not concerned with the question of the validity of Tolman's thermodynamics; and accordingly we make full use of his method of attacking the problem in hand.

---

* Communicated by the Author.

§2. DERIVATION OF THE EQUATION.

We adopt a spherically symmetric line element in the form

$$ds^2 = -\exp\{\lambda(r)\}\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2) + \exp\{\nu(r)\}\,dt^2. \quad (2.1)$$

In the following a number preceded by the letter T refers to the equations of the work of Tolman quoted above; and his notation is adhered to throughout. Then, by T(95.13),

$$8\pi p_0 = \exp\{-\lambda\}\left(\frac{\nu'}{r} + \frac{1}{r^2}\right) - \frac{1}{r^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

$$8\pi \rho_{00} = \exp\{-\lambda\}\left(\frac{\lambda'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

Consequently

$$\begin{aligned}\delta E_0 &= \delta(4\pi\rho_{00}\exp\{\tfrac{1}{2}\lambda\}r^2\,dr)\\ &= (-2\pi\rho_{00}r^2 + \tfrac{1}{2})\exp\{\tfrac{1}{2}\lambda\}\delta\lambda\,dr + \tfrac{1}{2}r\exp\{-\tfrac{1}{2}\lambda\}\delta\lambda'\,dr,\end{aligned}$$

by (2.3). Also, by (2.3),

$$p_0\delta v_0 = p_0\delta(4\pi\exp\{\tfrac{1}{2}\lambda\}r^2\,dr) = 2\pi p_0 r^2\exp\{\tfrac{1}{2}\lambda\}\delta\lambda\,dr.$$

Hence

$$\begin{aligned}\delta E_0 + p_0\delta v_0 &= [-2\pi(\rho_{00} - p_0)r^2 + \tfrac{1}{2}]\exp\{\tfrac{1}{2}\lambda\}\delta\lambda\,dr + \tfrac{1}{2}r\exp\{-\tfrac{1}{2}\lambda\}\delta\lambda'\,dr\\ &= \{[\tfrac{1}{4}r(\nu' - \lambda') + \tfrac{1}{2}]\delta\lambda + \tfrac{1}{2}r\delta\lambda'\}\exp\{-\tfrac{1}{2}\lambda\}\,dr. \quad . \quad . \quad . \quad . \quad (2.4)\end{aligned}$$

Inserting (2.4) in the condition for thermodynamic equilibrium T(124.5) we have

$$\int\frac{\exp\{-\frac{1}{2}\lambda\}}{2T_0}\{[\tfrac{1}{2}r(\nu' - \lambda') + 1]\delta\lambda + r\delta\lambda'\}\,dr = 0 \quad . \quad . \quad . \quad (2.5)$$

for all arbitrary variations which vanish on the boundary of the region considered. Integrating the term containing $\delta\lambda'$ by parts, and noting that the integrated part vanishes by hypothesis, (2.5) becomes

$$\begin{aligned}&\int\left\{\frac{\exp\{-\frac{1}{2}\lambda\}}{2T_0}[\tfrac{1}{2}r(\nu' - \lambda') + 1] - \left(\frac{r\exp\{-\frac{1}{2}\lambda\}}{2T_0}\right)'\right\}\delta\lambda\,dr\\ &\qquad = \tfrac{1}{2}\int r\exp\{-\tfrac{1}{2}\lambda\}\left[\frac{\nu'}{2T_0} - \left(\frac{1}{T_0}\right)'\right]\delta\lambda\,dr = 0.\end{aligned}$$

Accordingly we must have

$$\left(\frac{1}{T_0}\right)' = \frac{\nu'}{2T_0}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.6)$$

This is the first order equation which replaces T(128.2); and its integral is again given by (1.3).

REFERENCE.

TOLMAN, R. C., 1934, *Relativity, Thermodynamics and Cosmology*, (Oxford University Press), Ch. IX., § 128, pp. 312–314.

XXX. *Scattering of Pseudoscalar Charged Mesons by Nucleons.*—II.

By E. Corinaldesi* and G. Field†,
Dublin Institute for Advanced Studies‡.

[Received January 4, 1950.]

Summary.

In §1 the fourth- and sixth-order contributions to the scattering cross-section are evaluated for the non-relativistic region ($\omega \not> 3\,\mu$). The other section deals with damping effects.

§1. Sixth-order Corrections to the Scattering Cross-section.

We shall now investigate the same fourth-order transitions which were discussed in §5 of Part I. (Corinaldesi and Field 1949). A more refined, still non-relativistic, treatment will be given here, for

$$\frac{\mathrm{P}_i^2}{\mathrm{M}^2},\quad \frac{\mathrm{p}_i^2}{\mathrm{M}^2},\quad \frac{\mathrm{P}_f^2}{\mathrm{M}^2},\quad \frac{\mathrm{p}_f^2}{\mathrm{M}^2} \ll 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

but with

$$\mathrm{P}_i,\, \mathrm{p}_i,\, \mathrm{P}_f,\, \mathrm{p}_f \sim \mu, \qquad (\omega \not> 3\mu). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In order to simplify the calculation we shall choose as our reference system that in which the centre of gravity of the proton and meson is at rest. This will not affect the value which will be obtained for the cross-section. In the new reference system,

$$\left.\begin{aligned} \mathbf{P}_i &= -\mathbf{p}_i, \\ \mathbf{P}_f &= -\mathbf{p}_f. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

From the conservation of energy between the initial and final states, it follows that

$$|\mathbf{P}_i| = |\mathbf{p}_i| = |\mathbf{P}_f| = |\mathbf{p}_f| \;(=\mathrm{p},\text{ say}) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and

$$(\mathbf{p}_i \cdot \mathbf{p}_f) = \mathrm{p}^2 \cos\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where θ is the angle of scattering. We have also, with the notation of Part I.,

$$\omega_i = \omega_f = \omega$$

and

$$\Omega_i = \Omega_f = \Omega. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

---

* Now at Physics Department, University of Manchester.

† Now at Department of Mathematical Physics, University of Birmingham.

‡ Communicated by the Authors.

When conditions (1) are (2) are imposed, only eleven transitions are found to give contributions of the fourth-order which are not negligibly small. They are the same transitions that were discussed in Part I. All others have to be neglected, even for the higher energy range now under consideration.

Analogously to formula (49) of Part I. we have now

$$\mathrm{H}_{2,4}+\mathrm{H}_{4,2}=\frac{1}{8\Omega^2}\mathrm{Sp}\,\{\mathscr{H}_2(\Omega+\boldsymbol{\alpha}\,.\,\mathbf{P}_i+\mathrm{M}\beta)\mathscr{H}_4(\Omega+\boldsymbol{\alpha}\,.\,\mathbf{P}_f+\mathrm{M}\beta)$$
$$+\mathscr{H}_4(\Omega+\boldsymbol{\alpha}\,.\,\mathbf{P}_i+\mathrm{M}\beta)\mathscr{H}_2(\Omega+\boldsymbol{\alpha}\,.\,\mathbf{P}_f+\mathrm{M}\beta)\} \quad . \quad . \quad (7)$$

where

$$\mathscr{H}_2=-\frac{2\pi \mathrm{g}^2}{\mathrm{V}\omega}\left\{\frac{\omega-\Omega+\boldsymbol{\alpha}\,.\,\overline{\mathbf{P}}+\mathrm{M}\beta}{(\Omega-\omega)^2-\bar{\Omega}^2}\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\mathscr{H}_4=\frac{\pi^2\mathrm{g}^4}{2\mathrm{V}^2\omega}\sum_{\mathbf{p}'}\frac{1}{\omega(\mathbf{p}')\Omega(\mathbf{p}_1)\Omega(\mathbf{p}_2)\Omega(\mathbf{p}_3)(\mathrm{E}_i-\mathrm{E}_l)(\mathrm{E}_i-\mathrm{E}_m)(\mathrm{E}_i-\mathrm{E}_n)}\{\beta\gamma_5[\Omega(\mathbf{p}_1)$$
$$\mp\boldsymbol{\alpha}\,.\,\mathbf{p}_1\pm\mathrm{M}\beta]\beta\gamma_5[\Omega(\mathbf{p}_2)\pm\boldsymbol{\alpha}\,.\,\mathbf{p}_2\mp\mathrm{M}\beta]\beta\gamma_5[\Omega(\mathbf{p}_3)\mp\boldsymbol{\alpha}\,.\,\mathbf{p}_3\pm\mathrm{M}\beta]\beta\gamma_5\}. \quad . \quad . \quad . \quad (9)$$

In (9)

$$\left.\begin{aligned}\mathbf{p}_2&=\mathbf{p}',\\ \mathbf{p}_1&=\mathbf{p}'+\mathbf{p}_f \text{ or } -\overline{\mathbf{P}},\\ \mathbf{p}_3&=\mathbf{p}'+\mathbf{p}_i \text{ or } -\overline{\mathbf{P}}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Performing the integration over momentum space, and neglecting terms of second and higher orders in p/M as small compared with 1, we obtain

$$\mathrm{H}_{2,4}+\mathrm{H}_{4,2}=-\frac{3\pi}{4}\frac{\mathrm{g}^6}{\mathrm{V}^2\mathrm{M}^2\omega^2}\left(2+3\frac{\omega}{\mathrm{M}}-\frac{\mathrm{p}^2}{\mathrm{M}\omega}\cos\theta\right). \quad . \quad . \quad (11)$$

Similarly, $|\mathrm{H}_2|^2$ is evaluated from (8) and formula (16) of Part I.

$$|\mathrm{H}_2|^2=\frac{\pi^2\mathrm{g}^4}{\mathrm{V}^2\mathrm{M}^2\omega^2}\left(1+\frac{\omega}{\mathrm{M}}-\frac{\mathrm{p}^2}{\mathrm{M}\omega}\cos\theta\right). \quad . \quad . \quad . \quad . \quad (12)$$

The differential cross-section, dσ, is then given by

$$\begin{aligned}\mathrm{d}\sigma&=\mathrm{d}\sigma_4+\mathrm{d}\sigma_6+\ldots\\ &=\frac{\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi}{(2\pi)^2}\omega^2\mathrm{V}^2[|\mathrm{H}_2|^2+\mathrm{H}_{2,4}+\mathrm{H}_{4,2}+\ldots],\end{aligned} \quad . \quad . \quad . \quad (13)$$

in analogy with formula (55) of Part I. Thus the fourth-order term is

$$\mathrm{d}\sigma_4=\frac{\mathrm{g}^4}{4\mathrm{M}^2}\left(1+\frac{\omega}{\mathrm{M}}-\frac{\mathrm{p}^2}{\mathrm{M}\omega}\cos\theta\right)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \quad . \quad . \quad . \quad (14)$$

and the sixth-order term

$$\mathrm{d}\sigma_6=-\frac{3}{16\pi\mathrm{M}^2}\mathrm{g}^6\left(2+3\frac{\omega}{\mathrm{M}}-\frac{\mathrm{p}^2}{\mathrm{M}\omega}\cos\theta\right)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi, \quad . \quad (15)$$

where sin θ dθ dφ is the element of solid angle.

Integrating over θ and ϕ, we obtain for the fourth-order contribution to the *total* cross-section

$$\sigma_4 = \frac{\pi}{M^2} g^4 \left(1 + \frac{\omega}{M}\right) \quad . \; . \; . \; . \; . \; . \; . \; . \qquad (14')$$

and for the sixth-order contribution

$$\sigma_6 = -\frac{3}{4M^2} g^6 \left(2 + 3\frac{\omega}{M}\right). \quad . \; . \; . \; . \; . \; . \; . \qquad (15')$$

The ratio of the sixth- to the fourth-order terms is then given by

$$\left|\frac{\sigma_6}{\sigma_4}\right| = \frac{3}{4\pi} g^2 \left(2 + \frac{\omega}{M}\right). \quad . \; . \; . \; . \; . \; . \; . \; . \qquad (16)$$

This formula shows that, to the order of approximation used here, the ratio of the two contributions to the cross-section increases linearly with the energy. As for the convergence of the series $\sigma_4 + \sigma_6 + \ldots$, this depends, as was pointed out in Part I., on the choice of the numerical value of the coupling constant g.

At the present stage of the development of meson theory it is too early to say which of the values proposed by various authors is likely to be in agreement with experimental results—in so far as these already exist. In fact it is doubtful whether a pure pseudoscalar theory with a pseudoscalar coupling is likely to give a satisfactory account of the scattering process considered here. Several authors have adopted the value $g^2 \sim 36$, which would make the series divergent (Villars 1947, Dyson 1948, Luttinger 1948). Slotnick and Heitler (1949) have suggested a slightly lower value, $g^2 \sim 20$. On the other hand, Møller and Rosenfeld (1940), in their mixed theory, have chosen for the pseudoscalar coupling of pseudoscalar mesons a value $g^2 \sim 0{\cdot}1$, for which the series would converge. Recently, Bethe (1949) has been proposing yet another value, $g^2 \sim 4$.

Finally, we want to point out that the numerical value obtained for σ_6 depends on the way in which the renormalization of g is carried out. Since this point was not discussed in detail in Part I. of this paper, we shall give a short summary of the procedure here*:

After having carried out the mass and "charge" renormalization, we are left with $\sigma_6 = \sigma_6' + \sigma_6''$. Here, σ_6' is what remains of the previously divergent matrix elements, and σ_6'' is convergent. Now, σ_6' vanishes in the extreme non-relativistic case (E.N.R.) and even in the approximation (1) considered in this paper it does not give any contribution to the cross-section (15'), which is due only to σ_6''. In the case of the Compton effect, as treated by Corinaldesi and Jost (1948), the situation was different. There σ_6'', too, was zero in the E.N.R. case and this circumstance made it possible to say that the renormalization of the electric charge had been performed in such a way that $\lim \sigma_6=0$ for $p\to 0$. If we would apply the same criterion in the present calculation, the whole contribution to σ_6 which is constant for $p\to 0$ would have to be left out, and σ_6 would be much smaller than σ_4 for all energies under consideration. But we see no reason why we should apply this criterion to meson processes.

---

* We are grateful to Dr. M. P. Schafroth, Zürich, for asking us to clarify this point.

§ 2. Damping Effects.

Another correction to the cross-section may be obtained if we consider the reaction of the emitted meson on the motion of the nucleon. Attempts have been made by several authors to take account of this damping reaction (Heitler and Peng 1942, Wilson 1941, Sokolov 1941, Gora 1943) and we shall briefly discuss its effects on the results obtained above.

The scattering cross-section may be written in the form

$$\sigma_D=2\pi\rho(\omega)|U|^2, \qquad \ldots \ldots \ldots \quad (17)$$

where U is a probability amplitude defined by the integral equation

$$U=K-i\pi K\rho U, \qquad \ldots \ldots \ldots \quad (18)$$

U is a matrix element depending on the initial and final states, K is a similar matrix element given by $K=H_2+H_4+\ldots$ and $K\rho U$ is a matrix product, with summation over all final states. In order to obtain an idea of the order of magnitude of the damping effect it will not be necessary for us to solve the integral equation (18) rigorously. Substitution for U, in the damping term, of an average over angles will be sufficient for our purpose. We then obtain

$$U=\frac{K}{1+i\pi\rho K}, \qquad \ldots \ldots \ldots \quad (19)$$

so that the corrected cross-section σ_D is given by

$$\sigma_D=\frac{\sigma}{1+\kappa^2}, \qquad \ldots \ldots \ldots \quad (20)$$

where

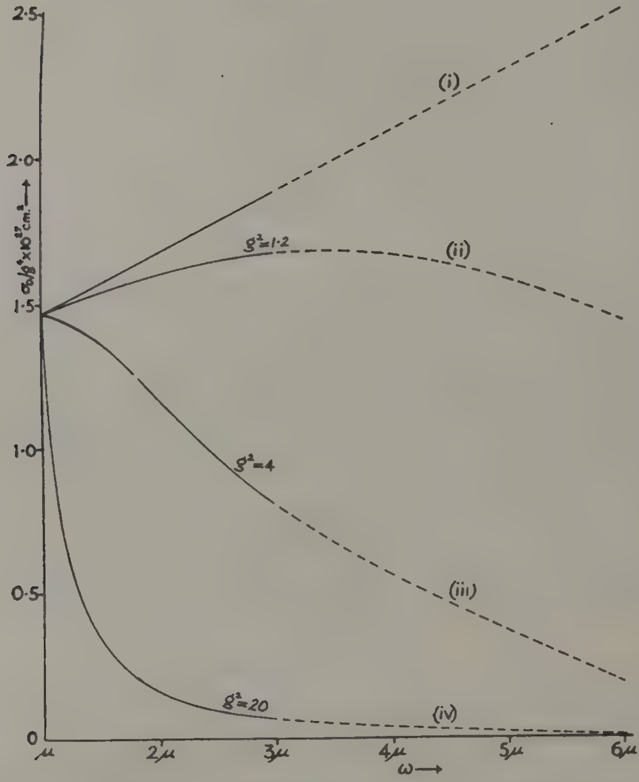
$$\sigma=2\pi\rho(\omega)|K|^2 \qquad \ldots \ldots \ldots \quad (21)$$

and

$$\kappa^2=\pi^2\rho^2|K|^2. \qquad \ldots \ldots \ldots \quad (22)$$

We shall now discuss the behaviour of σ_D for different values of g^2. For very high values of the coupling constant the series for $|K|^2$ diverges. The damping would still make the cross-section finite, namely $\sigma_D=\frac{2}{\pi\rho}$, but there is hardly any sense in including higher-order terms in the expressions for $|K|^2$ and κ^2, and this result would be very unreliable. In this case, we might rather assume that the first non-vanishing approximation of K comes closer to describing the real physical process than the inclusion of the rapidly increasing higher-order approximations would. Then we should have to put $|K|^2=|H_2|^2$. At very low energies, the values of σ_D for $g^2=20$ are, of course, very much higher than those for $g^2=4$. With increasing energy, however, damping is much greater in the former case (since $\kappa^2 \propto g^4$) and at the upper limit of the range for which the formulæ are still valid (see (2)), cross-sections for both values of g^2 are of the same order of magnitude ($\sim 2\times 10^{-26}$ cm.^{2}, see fig. 1).

Fig. 1.

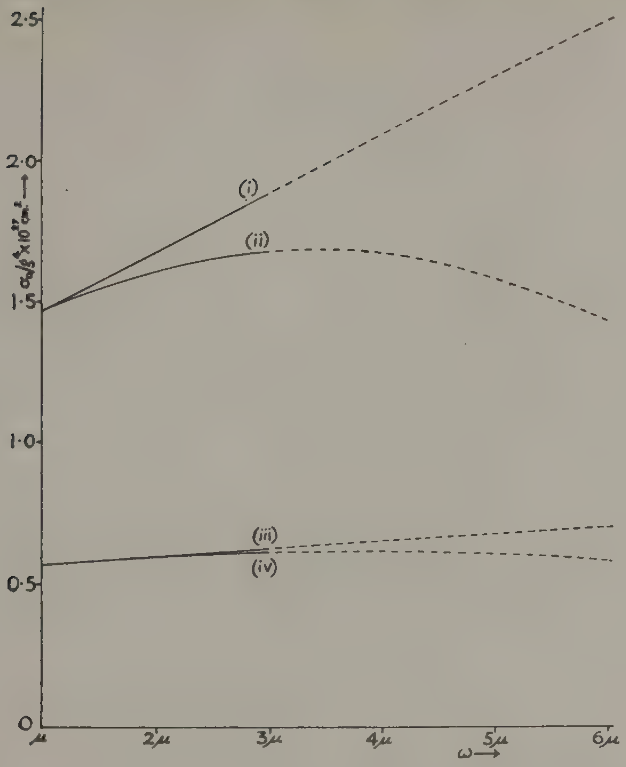


Energy dependence of damped cross-section, with $|K|^2=|H_2|^2$.
(i.) σ/g^4 (no damping),
(ii.), (iii.) and (iv.) σ_D/g^4 for $g^2=1\cdot2$, 4 and 20 respectively.

For lower values of g^2, *e. g.* $g^2=1\cdot2$, the damping effect is much smaller than in the previous cases. Moreover, it is now reasonable to include the sixth-order terms in the expression for $|K|^2$. Fig. 2 shows the damping effect for the two cases $|K|^2=|H_2|^2$ and $|K|^2=|H_2|^2+H_{2,4}+H_{4,2}$.

In conclusion, we would like to thank Professor W. Heitler for much valuable criticism.

Fig. 2.



Influence of sixth-order correction on cross-section, for $g^2=1{\cdot}2$.

(i.) σ/g^4 (no damping), with $|K|^2=|H_2|^2$.
(ii.) σ_D/g^4 with $|K|^2=|H_2|^2$.
(iii.) σ/g^4 (no damping), with $|K|^2=|H_2|^2+H_{2,4}+H_{4,2}$.
(iv.) σ_D/g^4 with $|K|^2=|H_2|^2+H_{2,4}+H_{4,2}$.

REFERENCES.

BETHE, H. A., 1949, *Phys. Rev.*, **76,** 191.
CORINALDESI, E., and FIELD, G., 1949, *Phil. Mag.*, **40,** 1159.
CORINALDESI, E., and JOST, R., 1948, *Helv. Phys. Acta*, **21,** 183.
DYSON, F. J., 1948, *Phys. Rev.*, **73,** 929.
GORA, E., 1943, *Zeits. f. Physik.*, **120,** 121.
HEITLER, W., and PENG, H. W., 1942, *Proc. Camb. Phil. Soc.*, **38,** 296.
LUTTINGER, J. M., 1948, *Helv. Phys. Acta*, **21,** 483.
MØLLER, C., and ROSENFELD, L., 1940, *D. Kgl. Danske Vid. Selsk.*, **17,** No. 8.
SLOTNICK, M., and HEITLER, W., 1949, *Phys. Rev.*, **75,** 1645.
SOKOLOV, A., 1941, *J. Phys. U.S.S.R.*, **5,** 231.
VILLARS, F., 1947, *Helv. Phys. Acta*, **20,** 476.
WILSON, A. H., 1941, *Proc. Camb. Phil. Soc.*, **37,** 301.

XXXI. *Influence of the Overlap Integral in Molecular-orbital Calculations for Alternant Hydrocarbons.*

By J. de Heer*,
Wheatstone Physics Department, King's College, London †.

[Received January 10, 1950].

Summary.

Some recent studies have indicated that in molecular-orbital calculations on alternant hydrocarbons the explicit inclusion of the overlap integral, S, between adjacent atomic orbitals makes no significant difference in the consistency of the results (Wheland 1942, Brooks 1941, Coulson and Chirgwin, in press). In the present paper a further investigation is reported on total π-electronic energies and the various polarizabilities of such hydrocarbons. In connection with a regularity observed by Wheland (1941), the resonance energies are discussed in relation to the total π-electronic energies; it appears that the latter are more consistent than the former. A correlation between the atom–atom–" self-polarizabilities " and the conjugation energies across single bonds leads us to define " conjugating powers " of hydrocarbon residues in complete concordance with the analogous definitions given in the simpler theory, neglecting overlap.

Introduction.

One of the most surprising features of the molecular-orbital method, if applied to ***alternant hydrocarbons*** (for a definition, see Coulson and Longuet-Higgins 1947 b), is the fact that the explicit inclusion of the overlap integral, S, between adjacent carbon $2p_z$-atomic orbitals makes no significant difference in the self-consistency of the calculations of several molecular properties. This was first shown by Wheland (1941) for the *resonance energies*, by comparing the results of calculations neglecting and including S respectively. In the same year Brooks (1941) found that computed values for the *diamagnetic anisotropy* of several aromatic ring systems are nearly unaffected by the inclusion of S. Recently, B. H. Chirgwin and C. A. Coulson (in press), in a more elaborate analysis, have proved that *mobile bond orders* and *π-electronic charge densities* remain completely unchanged, while they concluded that the various *self- and mutual polarizabilities* (introduced by Coulson and Longuet-Higgins 1947 a) are only slightly affected. Finally, unpublished calculations by C. A. Coulson indicate, that for a large number of polynuclear aromatic hydrocarbons the quotient β/γ, γ and β being the parameters in which the *total π-electronic energies* are expressed, when including and neglecting S respectively, is a fairly constant number. The last regularity is of the same character as—and obviously related to—the one concerning resonance energies.

---

* Now at the department of general and inorganic chemistry, the University of Amsterdam.

† Communicated by Professor C. A. Coulson, D.Sc.

Only in the case of bond orders and charge densities has a theoretical explanation been given. In all other cases the problem is strikingly similar: The quantities considered are the sum of certain contributions from the individual molecular orbitals or from pairs of them (*e. g.* the total π-electronic energy is twice the sum of the energies of the individual m.o.'s and the various polarizabilities are built up from contributions of pairs of one occupied and one unoccupied m.o.). Now it appears that while these separate contributions, being affected by some factor(s) containing S, may and in most cases do change appreciably and irregularly, their sum for some reason to be given either changes very little or is merely multiplied by a number, which is approximately the same for all molecules.

In the present investigation the explanation will be proved to lie in the fact that S, the value of which is usually taken to be 0·25, only occurs in the summed expressions as S^2, which is obviously a small quantity. First the total π-electronic energies and the resonance energies and then the polarizabilities will be discussed (the diamagnetic anisotropy is at present not tractable by our methods). It will be shown that total π-electronic energies are more consistent than resonance energies. Finally, conjugation energies, conjugating powers and their interrelation (Coulson and Longuet-Higgins 1948) will also be considered with the same inclusion of overlap. It turns out that here again certain conclusions drawn from the simpler theory, neglecting S, remain valid.

While the present analysis together with previous investigations referred to above, justifies us in neglecting overlap in molecular-orbital calculations on the ground-state of aromatic and olefinic hydrocarbons, this is no longer true in the case of heteromolecules. However, it still seems reasonable to assume that if few and small hetero-parameters occur in our equations, the inclusion of S will not disturb the consistency of the results essentially. If many and large such parameters are to be introduced the value of the entire simple molecular-orbital method as such becomes very doubtful.

1. The Total π-Electronic Energy.

If we neglect S, we can write for the energy of the r^{th} m.o., ϵ_r,

$$\epsilon_r = k_r \beta. \qquad (1)$$

Consequently the total π-electronic energy is given by

$$\mathscr{E} = \sum_{j=1}^{m} 2\epsilon_r, \qquad (2)$$

where the summation index, j, runs over the m occupied orbitals.

If we include S, taken to have the same numerical value for all pairs of adjacent atoms, we can write for the energy of the r^{th} m.o., E_r,

$$E_r=m_r\gamma, \tag{3}$$

where the new energy-parameter, γ, is given by

$$\gamma=\beta-S\alpha, \tag{4}$$

The relation between m_r and k_r is known to be (Wheland 1941, Coulson and Chirgwin, in press):

$$m_r=\frac{k_r}{1+k_rS}. \tag{5}$$

The total π-electronic energy now becomes

$$\mathscr{E}_S=\sum_{j=1}^{m} 2E_r. \tag{6}$$

Our notation implies that α, the Coulomb integral for the carbon centres, is taken as the zero level from which all energies are measured. In addition it should be remembered that S is a positive quantity, but β and γ are negative, while k and m are positive for all bonding orbitals.

From (6), (3) and (1) we derive

$$\mathscr{E}_S=2\gamma\sum_{j=1}^{m}\frac{\epsilon_r}{\beta+S\epsilon_r}. \tag{7}$$

This sum can be expressed by the complex integral:

$$\mathscr{E}_S=\frac{\gamma}{\pi i}\int_{\Gamma}\left[\frac{z}{\beta+Sz}\cdot\frac{\Delta'(z)}{\Delta(z)}\right]dz, \tag{8}$$

where the path of integration, Γ, is the imaginary axis from $-\infty i$ to $+\infty i$ and the infinite semi-circle to the left of the y-axis (fig. 1; compare Coulson 1940, Coulson and Longuet-Higgins 1947 a). $\Delta(z)$ is the secular-determinant:

$$\Delta(z)=\begin{vmatrix} -z & \beta_{12} & \beta_{13}\cdots\cdots\beta_{1n} \\ \beta_{21} & -z & \beta_{23}\cdots\cdots\beta_{2n} \\ \vdots & & \\ \beta_{n1} & & -z \end{vmatrix} \tag{9}$$

Concerning equation (8) we first notice that the pole $z=-\beta/S$ lies outside the contour. Further, if we compare this result with the alternative expression given by Coulson and Chirgwin (in press; see equation (38) of their paper), then we observe that in (8) we have got rid of all quantities involving Coulson and Chirgwin's mixed tensor H^{ν}_{μ} and instead can deal with the same secular-determinant, (9), as in the simpler theory, neglecting S. This is achieved " at the cost of " introducing the complicating factor $1/(\beta+Sz)$ in the integrand.

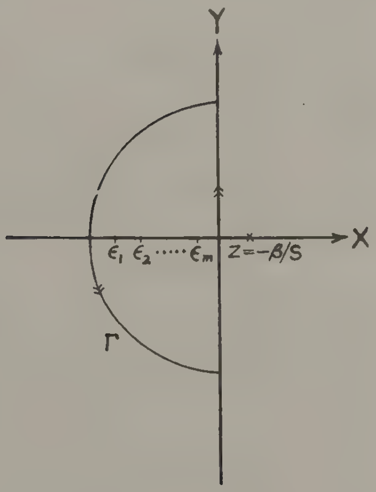
If we wish, we may replace (8) by

$$\mathscr{E}_S = \frac{\gamma}{\pi i}\int_\Gamma \left[\frac{z}{\beta+Sz}\cdot\frac{\Delta'(z)}{\Delta(z)} - \frac{n}{\beta+Sz}\right]dz, \quad . \quad . \quad . \quad . \quad (10)$$

where n is the number of atoms in the molecule under consideration. Replacing (8) by (10) is correct, because

$$\int_\Gamma \frac{dz}{\beta+Sz} = 0,$$

Fig. 1.



and it has the advantage that the contribution over the infinite semi-circle vanishes, since for large z,

$$\frac{z}{\beta+Sz}\cdot\frac{\Delta'(z)}{\Delta(z)} - \frac{n}{\beta+Sz} \cong \frac{1}{S}\cdot\frac{nz^{n-1}}{z^n} - \frac{n}{Sz} = 0.$$

Thus the integral (10) reduces to

$$\mathscr{E}_S = \frac{\gamma}{\pi}\int_{-\infty}^{\infty} \left[\frac{iy}{\beta+Siy}\cdot\frac{\Delta'(iy)}{\Delta(iy)} - \frac{n}{\beta+Siy}\right]dy. \quad . \quad . \quad . \quad (11)$$

Multiplying numerator and denominator of the integrand by $(\beta - Siy)$ yields

$$\mathscr{E}_S = \frac{\gamma}{\pi}\int_{-\infty}^{\infty} \left[\frac{\beta iy+Sy^2}{\beta^2+S^2y^2}\cdot\frac{\Delta'(iy)}{\Delta(iy)} + \frac{-n\beta+nSiy}{\beta^2+S^2y^2}\right]dy. \quad . \quad . \quad (12)$$

As $\frac{\Delta'(iy)}{\Delta(iy)}$ is an odd function of y, while of course $(\beta^2+S^2y^2)$ is even, the terms

$$\frac{Sy^2}{\beta^2+S^2y^2}\cdot\frac{\Delta'(iy)}{\Delta(iy)} \quad \text{and} \quad \frac{nSiy}{\beta^2+S^2y^2}$$

do not contribute to the integral. Hence equation (12) reduces to

$$\mathscr{E}_S=\frac{\gamma}{\pi}\int_{-\infty}^{\infty}\left[\frac{\beta iy}{\beta^2+S^2y^2}\cdot\frac{\Delta'(iy)}{\Delta(iy)}-\frac{n\beta}{\beta^2+S^2y^2}\right]dy,$$

which we finally write as

$$\mathscr{E}_S=\frac{\gamma}{\beta}\cdot\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{\{1+S^2(y/\beta)^2\}}\cdot\left\{iy\frac{\Delta'(iy)}{\Delta(iy)}-n\right\}dy. \quad . \quad . \quad (13)$$

If we put S equal to zero, we obtain

$$\mathscr{E}=\frac{1}{\pi}\int_{-\infty}^{\infty}\left\{iy\frac{\Delta'(iy)}{\Delta(iy)}-n\right\}dy, \quad . \quad . \quad . \quad . \quad . \quad (14)$$

which is equivalent to the result of the simpler theory (Coulson and Longuet-Higgins 1947 a), provided again that all energies are measured from α.

It will be noticed, that in (13) S only occurs as the small quantity S^2. Furthermore Coulson has shown (1940), that only the interval

$$0\leqslant|(y/\beta)|\leqslant 4$$

need be considered, in which interval the " correction-factor " varies from 1 to $\frac{1}{2}$. Finally the integrand in (14) usually reaches a rather steep maximum in the neighbourhood of $|y/\beta|=2$ (Coulson, private communication). If this maximum alone determined $\mathscr{E}_S$ in (13), then we should have

$$\mathscr{E}_S=0{\cdot}8\frac{\gamma}{\beta}\mathscr{E},$$

and we can expect the proper result to be only slightly different. In fact, as we shall see in the next paragraph, if we equate $\mathscr{E}$ and $\mathscr{E}_S$ as obtained by direct calculations, the result for a large number of molecules is a value for β/γ of about 0·70–0·75.

2. Resonance Energies and Total π-Electronic Energies; the Quotient β/γ.

In this paragraph $(\beta/\gamma)_{\mathscr{E}}$ stands for the quotient β/γ which is found by equating $\mathscr{E}$ and $\mathscr{E}_S$, while $(\beta/\gamma)_R$ stands for the same quotient when obtained by equating the corresponding resonance energies, R and R_S. In Table I. values for $(\beta/\gamma)_{\mathscr{E}}$ and $(\beta/\gamma)_R$ for a number of molecules and free radicals have been collected. Most data for calculating these values have been published (see *e. g.* Wheland 1941, who used β and γ in the reverse sense); a few new computations were necessary.

As already mentioned in the introduction, Wheland (1941) pointed out that $(\beta/\gamma)_R$ is fairly constant. However, on closer inspection, it cannot be denied that rather large differences occur (up to 30 per cent from the mean value, 0·50 !), in particular between different types of molecule. On the other hand a glance at the last column of Table I. shows the values of $(\beta/\gamma)_{\mathscr{E}}$, which were not considered hitherto, to be far more consistent (0·73 ± 3 per cent, if for the moment we consider ethylene as a separate case). This last fact is in accordance with the theoretical ideas developed in paragraph 1.

TABLE I.

Molecule	$(\beta/\gamma)_R$	$(\beta/\gamma)_{\mathscr{E}}$	
Benzene	0·54	0·73	
Biphenyl	0·51	0·72	
Naphthalene	0·51	0·72	
Anthracene	0·49	0·72	
Phenanthrene	0·50	0·72	
Pyrene	0·49	0·71	
1, 3, 5-triphenylbenzene	0·50	0·72	
Fulvene	0·44	0·73	*
Azulene	0·48	0·72	
Styrene	0·50	0·73	
Stilbene	0·50	0·72	
1, 1-diphenyl-ethylene	0·50	0·72	
1-phenyl-butadiene	0·48	0·73	
2-phenyl-butadiene	0·48	0·73	
Ethylene	—	0·80	
Butadiene	0·36	0·75	
Hexatriene	0·39	0·74	
Octatetræne	0·40	0·74	
Decapentæne	0·41	0·73	
Dodecahexæne	0·42	0·73	
Benzyl	0·56	0·72	
α-Naphthylmethyl	0·65	0·72	
β-Naphthylmethyl	0·64	0·72	
m-biphenylmethyl	0·56	0·72	
p-biphenylmethyl	0·58	0·72	
diphenylmethyl	0·53	0·72	
triphenylmethyl	0·51	0·72	
Allyl	0·63	0·75	

---

* These hydrocarbons contain odd-membered rings, and are therefore not alternant.

It can easily be shown, that to a first approximation, a constant value of $(\beta/\gamma)_{\mathscr{E}}$ of about 0·73 need not give rise at all to a similar uniformity amongst $(\beta/\gamma)_R$-values. Let us take the case of an alternant hydrocarbon with an even number, n, of carbon atoms. Let us put

$$\mathscr{E} = K\beta.$$

Then according to our assumption above,

$$\mathscr{E}_S \cong 0{\cdot}73K\gamma.$$

To obtain R and R_S in the usual way, we have to subtract from these total π-electronic energies $n/2$ times the total π-electronic energy of an ethylene molecule (2β and $1{\cdot}6\gamma$ respectively). Hence

$$R = (K - n)\beta,$$

$$R_S \cong (0{\cdot}73K - 0{\cdot}8n)\gamma.$$

Equating R and R_S gives us the result

$$(\beta/\gamma)_R \cong 0{\cdot}73 - 0{\cdot}07 \cdot \frac{1}{\left\{\frac{K}{n} - 1\right\}}. \qquad \dots\dots (15)$$

As K/n is greater than 1, $(\beta/\gamma)_R$ can be expected to be smaller than 0·73. This indeed is found by direct calculations (Table I.). Furthermore, as K/n may show rather large fluctuations, equation (15) cannot be expected to yield very similar values of $(\beta/\gamma)_R$ for all types of molecule; *e. g.* comparing benzene with the open chain hydrocarbon with the same n, hexatriene, shows us very clearly what happens. For these molecules $(\beta/\gamma)_{\mathscr{E}}$ is 0·73 and 0·74 respectively. For benzene $K/n = 1{\cdot}33$, for hexatriene $K/n = 1{\cdot}16$. As a result we find a large difference in $(\beta/\gamma)_R$ (the values are 0·54 and 0·39 respectively). In this way, a detailed analysis of K/n might give us, for some type of molecule or radical, a more complete explanation of the $(\beta/\gamma)_R$ values as obtained by direct calculations, but this seems of minor importance to us.

Obviously the cause of all these difficulties lies in the fact that ethylene, which plays such an important part in the calculation of the resonance energies, stands alone in having an exceptionally high value of $(\beta/\gamma)_{\mathscr{E}}$, namely 0·8. There is no way of getting round this difficulty and the important point arising from this discussion is that resonance energies calculated with the inclusion of overlap for different types of molecule need not be reflected very accurately by the values obtained by means of the simpler theory, neglecting S. Total π-electronic energies appear to be far more consistent and should be given preference if such is possible within the scope of a particular investigation.

3. POLARIZABILITIES.

The definition of the various polarizabilities and several formulæ to calculate them have been given by Coulson and Longuet-Higgins (1947 a) within the framework of the simpler theory, neglecting overlap. Coulson and Chirgwin (in press) have recently shown how the inclusion of S necessitates a change in the formulæ, while some numerical calculations indicate that apart from a factor β/γ there is extremely little difference between the two sets of figures thus obtainable.

The short analysis to follow will be given for the specific case of the atom–atom " self-polarizabilities ", but it holds equally well for all the others. When S is neglected we shall use the customary notation $\pi_{r,\,r}$ for the self-polarizability of carbon centre r in a particular alternant hydrocarbon ; when S is taken into account, we shall write $(\pi_{r,\,r})_{\mathrm{S}}$ instead. In Table II. a few calculated values have been collected to illustrate the similarity between the two cases, for whose existence we must find some explanation.

TABLE II.

Hydrocarbon residue	$\beta\,.\,\pi_{r,r}$	$\gamma\,.\,(\pi_{r,r})_{\mathrm{S}}$
Phenyl	0·398	0·359
2-Butadienyl	0·402	0·377
Vinyl	0·500	0·469
1-Butadienyl	0·626	0·601
1-Hexatrienyl	0·685	0·661

Without including S, we have (Coulson and Longuet-Higgins, 1947 a) :

$$\pi_{r,\,r}=4\sum_{j=1}^{m}\sum_{h=m+1}^{n}\frac{c_{rj}^{2}\,.\,c_{rh}^{2}}{\epsilon_j-\epsilon_h}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where ϵ_j stands for the energy of the j^{th} molecular orbital, as before, and c_{rj} is the coefficient of the r^{th} atomic orbital in this m.o. The summation indices j and h refer to the occupied and unoccupied molecular orbitals respectively. (In (16) the index h is used instead of the more customary index k ; this is done to prevent confusion with the k, introduced in equation (1).)

When including S, equation (16) has to be replaced by (Coulson and Chirgwin, in press)

$$(\pi_{r,\,r})_{\mathrm{S}}=\frac{4\beta}{\gamma}\sum_{j=1}^{m}\sum_{h=m+1}^{n}\frac{c_{rj}^{2}\,.\,c_{rh}^{2}}{\lambda_j^2\lambda_h^2(\epsilon_j-\epsilon_h)}, \quad . \quad . \quad . \quad . \quad (17)$$

where

$$\left.\begin{aligned}\lambda_j^2&=1/(1+k_j\mathrm{S})\\ \lambda_h^2&=1/(1+k_h\mathrm{S})\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and where k_j and k_h are related to ϵ_j and ϵ_h according to equation (1). Substituting (18) in (17) yields

$$(\pi_{r,\,r})_{\mathrm{S}}=\frac{4\beta}{\gamma}\sum_{j=1}^{m}\sum_{h=m+1}^{n}[1+(k_j+k_h)\mathrm{S}+k_jk_h\mathrm{S}^2]\frac{c_{rj}^{2}\,.\,c_{rh}^{2}}{\epsilon_j-\epsilon_h}.$$

The double sum in the r.h.s. consists of three terms of which the first one is evidently $\beta/\gamma\,.\,\pi_{r,\,r}$. The second term, the one involving S to the first degree, vanishes owing to the symmetric distribution of occupied and unoccupied m.o.'s around the zero-level (Coulson and Longuet-Higgins, 1947 b). So finally we obtain

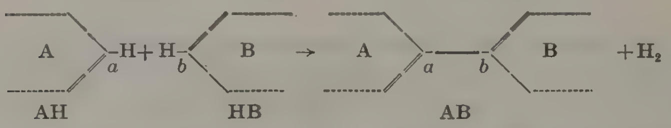
$$\gamma\,.\,(\pi_{r,r})_S=\beta\,.\,\pi_{r,r}+4S^2\sum_{j=1}^{m}\sum_{h=m+1}^{n}\frac{k_jk_h}{k_j-k_h}\,.\,c_{rj}^2c_{rh}^2. \quad \ldots \quad (19)$$

So the difference between $\gamma\,.\,(\pi_{r,\,r})_S$ and $\beta\,.\,\pi_{r,\,r}$ will always be small as it only involves S^2. Furthermore this " correction-term " will always be negative (compare Table II.) as all terms in the sum in the r.h.s. of (19) are negative. If we wish, we can use equation (19) for the numerical calculation of $(\pi_{r,\,r})_S$ if $\pi_{r,\,r}$ is known. *E. g.* for ethylene (one occupied and one unoccupied m.o.), with S taken to be 0·25 as usual, we obtain immediately

$$\gamma(\pi_{r,r})_S=0{\cdot}500+4\,.\,\frac{1}{16}\cdot\frac{1\,.\,(-1)}{1-(-1)}\cdot\frac{1}{2}\cdot\frac{1}{2}=0{\cdot}500-\frac{1}{32}=0{\cdot}469,$$

a result which, of course, is also obtained by the direct application of equations (17) and (18).

4. CONJUGATION ENERGY AND CONJUGATING POWERS.



The " Conjugation energy across a single bond " was defined by Coulson and Longuet-Higgins (1948) as follows (see diagram above): " Let us connect two alternant hydrocarbons, AH and HB, by eliminating two hydrogen atoms at carbon atoms *a* of AH and *b* of HB respectively and subsequently forming the bond *a–b*. Then the conjugation energy across this bond, R_{a-b}, is defined by (see diagram above)

$$-R_{ab}=\mathscr{E}(AB)-\mathscr{E}(AH)-\mathscr{E}(HB), \quad . \quad . \quad . \quad . \quad (20)$$

where, in accordance with our previous notation, the $\mathscr{E}$'s refer to the total π-electronic energies of the species indicated in brackets ".

Coulson and Longuet-Higgins (1948) show that the following relation holds in good approximation :

$$-R_{ab}/\beta=\text{constant}\times\beta\,.\,\sqrt{(\pi_{a,a}\cdot\pi_{b,b})}, \quad . \quad . \quad . \quad . \quad (21)$$

where $\pi_{a,\,a}$ and $\pi_{b,\,b}$ are the self-polarizabilities of the carbon atoms *a* and *b* in the systems AH and HB respectively. On account of this result the self-polarizability is considered as the " conjugating power " of a hydrocarbon residue.

If we include the overlap integral S in our calculations, we must obviously replace the defining equation (20) by

$$-(\mathrm{R}_{ab})_{\mathrm{S}}=\mathscr{E}_{\mathrm{S}}(\mathrm{AB})-\mathscr{E}_{\mathrm{S}}(\mathrm{AH})-\mathscr{E}_{\mathrm{S}}(\mathrm{HB}). \quad . \quad . \quad . \quad (20\mathrm{a})$$

Fig. 2 (*a*).

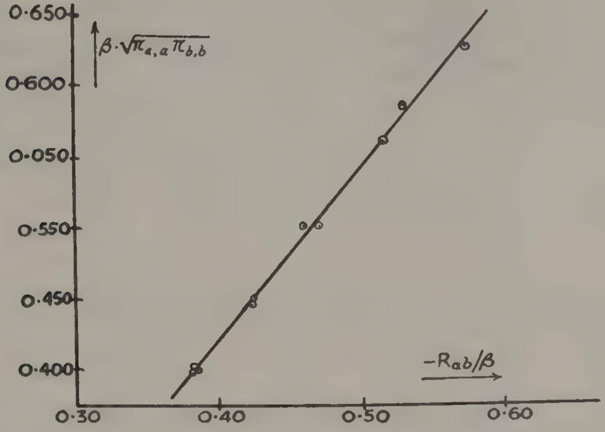
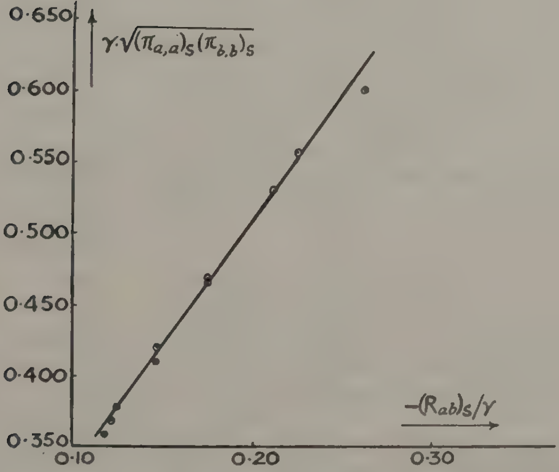


Fig. 2 (*b*).



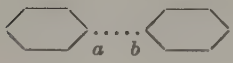
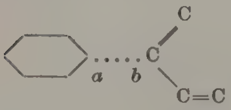
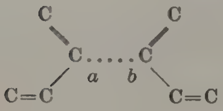
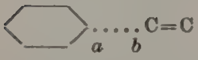
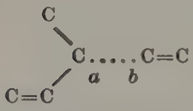
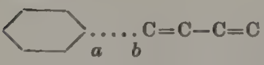
Then it is a natural question to ask whether a relationship still holds equivalent to the one embodied in (21), namely

$$-(\mathrm{R}_{ab})_{\mathrm{S}}/\gamma=\text{constant}\times\gamma\,.\,\sqrt{[(\pi_{a,a})_{\mathrm{S}}\cdot(\pi_{b,\,b})_{\mathrm{S}}]}. \quad . \quad . \quad (21\mathrm{a})$$

If so, then $(\pi_{a,a})_{\mathrm{S}}$ and $(\pi_{b,b})_{\mathrm{S}}$ act as conjugating powers, just as $\pi_{a,a}$ and $\pi_{b,b}$ did before.

A theoretical treatment is not possible here. It must be remembered that equation (21) was never derived rigorously, but even the equivalent to the approximate treatment, as given by Coulson and Longuet-Higgins (1948) cannot be given here, for the integral expressions for $\mathscr{E}_S$ and the $(\pi_{r,r})_S$'s become too involved for this purpose. So we can only check equation (21 a) by direct numerical calculations. In fig. 2 (*a*) we have therefore plotted computed values of $-R_{ab}/\beta$ against $\beta \,.\, \sqrt{[\pi_{a,a} \cdot \pi_{b,b}]}$ and in fig. 2 (*b*) $-(R_{ab})_S/\gamma$ against $\gamma \,.\, \sqrt{[(\pi_{a,a})_S \cdot (\pi_{b,b})_S]}$. The numerical data concerned have been collected in Table III.

TABLE III.

	$-R_{ab}/\beta$	$\beta . \sqrt{(\pi_{a,a} . \pi_{b,b})}$	$-(R_{ab})_S/\gamma$	$\gamma . \sqrt{[(\pi_{a,a})_S . (\pi_{b,b})_S]}$
	0·383	0·398	0·118	0·359
	0·386	0·400	0·121	0·368
	0·388	0·402	0·124	0·377
	0·424	0·446	0·146	0·410
	0·427	0·448	0·147	0·421
	0·460	0·499	0·174	0·464
C=C.....C=C a b	0·472	0·500	0·175	0·469
C=C−C=C.....C=C a b	0·516	0·559	0·212	0·531
C=C−C=C−C=C.....C=C a b	0·530	0·586	0·225	0·557
C=C−C=C.....C=C−C=C a b	0·574	0·626	0·263	0·601

A comparison of figs. 2 (*a*) and 2 (*b*) shows very clearly that equation (21 a) holds in as good an approximation as equation (21). This is not a complete surprise as we have shown in previous paragraphs that the values of $\mathscr{E}$ and $\pi_{r,r}$ are, in general, consistent with $\mathscr{E}_S$ and $(\pi_{r,r})_S$. These quantities completely determine the conjugation energies and the conjugating powers. Thus here again the explicit inclusion of the overlap integral does not upset the qualitative and semi-quantitative picture as given by the simpler theory. This result is particularly gratifying since some recent investigations show (Coulson, Evans and de Heer, to be published) that the conjugating powers in alternant hydrocarbons are also of significance in determining certain properties of the π-electronic system of substituted hydrocarbons.

Note added in proof.—Professor J. A. A. Ketelaar and Mr. J. van Dranen have drawn the attention of the writer to the following point: It is not clear from the manuscript that the complicated analysis of paragraph 1 is superfluous if our object is simply to conclude that total π-electronic energies are more consistent with respect to the inclusion of overlap than resonance energies. They point out that we have, by definition (notation as in paragraph 2):

$$\frac{\mathscr{E}_S}{\mathscr{E}}=(\beta/\gamma)_{\mathscr{E}}=\frac{0{\cdot}8n+R_S}{1{\cdot}0n+R}.$$

Combining the two facts that:

(i) R_S is small compared with $0{\cdot}8n$ and R is small compared with $1{\cdot}0n$,

(ii) Wheland (1941) has shown by direct calculations that R_S/R values are *fairly* constant,

it is more or less trivial that $\mathscr{E}_S/\mathscr{E}$ and $(\beta/\gamma)_{\mathscr{E}}$ are better constants than R_S/R and $(\beta/\gamma)_R$ respectively. It is also immediately clear from the above equation that the $(\beta/\gamma)_{\mathscr{E}}$ will in general be slightly smaller than 0·8 and can at most be equal to this value, namely if $R_S=R=0$ (ethylene).

In our investigation we have proved theoretically that the consistency of $\mathscr{E}_S/\mathscr{E}$—and thus of $(\beta/\gamma)_{\mathscr{E}}$—values and their magnitude ($\pm 0{\cdot}8$) can be expected *a priori*. We have subsequently shown, that if we take this result as a starting point and then proceed to define resonance energies in the usual way, it becomes immediately evident that $(\beta/\gamma)_R$-values will in general not show the same consistency. This second step is entirely equivalent to the analysis, given by Ketelaar and van Dranen (see above), but in the reverse sense.

ACKNOWLEDGMENT.

The writer wishes to thank Professor C. A. Coulson for many helpful suggestions and discussions. He also wishes to express his gratitude to Professor C. A. Coulson and Mr. B. H. Chirgwin for enabling him to read their manuscript prior to publication.

REFERENCES.

HARVEY BROOKS, 1941, *J. Chem. Phys.*, **9**, 463.
COULSON, C. A., 1940, *Proc. Camb. Phil. Soc.*, **36**, 201.
COULSON, C. A. and CHIRGWIN, B. H., *Proc. Roy. Soc.* (in press).
COULSON, C. A. and LONGUET-HIGGINS, H. C., 1947 a, *Proc. Roy. Soc.* A, **191**, 39; *Ibid.*, 1947 b, A, **192**, 16; *Ibid.*, 1948, A, **195**, 188.
WHELAND, G. W., 1941, *Journ. Am. Chem. Soc.*, **63**, 463.

XXXII. *The Estimation of Velocities and Local Mach Numbers in the Flow of a Compressible Fluid through Divergent Nozzles, under Adiabatic and Isothermal Conditions.*

By R. K. Tempest *.

[Received January 25, 1950.]

Summary.

In the present paper, a theory is developed giving the approximate velocity at any point in an inviscid compressible fluid flowing through a divergent nozzle. The method of calculation assumes a radial pattern for the stream lines, as suggested by the theoretical and experimental work on flows in diffusers carried out by Gourzhienko (1939). The velocity field is extended by successive applications of the known solution for radial flow from a point source. The formulæ obtained in the present paper are applicable to either subsonic or supersonic flows, they may be applied over any length of tube ; and from them the variation of the local speed of sound and Mach number may be estimated.

Numerical examples are given in tables to demonstrate the very large increases in local velocity and Mach number in supersonic flows associated with comparatively small variation of the expansion ratio of the nozzles.

A comparison with the results of one-dimensional theory has been made in the case where the velocity is approximately uniform across sections perpendicular to the axis of the nozzle. It is pointed out that one-dimensional theory may not be applied to a flow in which a considerable change of velocity occurs across a section ; this type of flow is described by the theory of this paper.

§ 1. Introduction.

We consider the flow of a frictionless, compressible fluid in a nozzle with axial symmetry. In general discussions of the flow of compressible fluids in tubes, linearized equations of motion are used to provide a solution valid in some small length of tube. The assumption is made that the Mach number and the local speed of sound are constant over this length, or, more accurately, variations in these quantities are assumed to be comparatively small and are neglected in the coefficients of terms that are themselves small. Numerical evaluation of the velocity at any point is cumbersome owing to the complexity of the exact differential equation for the potential governing the flow. Under certain circumstances, however, these equations may be simplified to give a formula for the velocity at any point in very simple form. The theory given below has the advantages that it may be applied over any length of tube and that the variation of the speed of sound and local Mach number are taken into consideration and estimated.

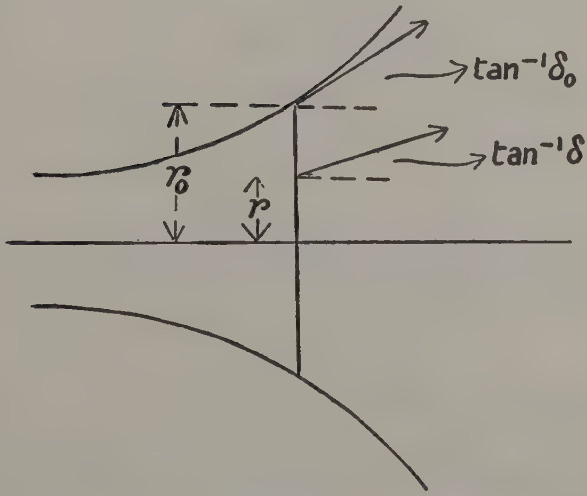
---

* Communicated by the Author.

We shall first investigate the flow in a cone which is purely radial from a source at the vertex and in which the slope of a stream line through any point in the field is proportional to the displacement of the point from the central stream line. This state of affairs gave Gourzhienko good experimental agreement in the case of flows which are not completely conical and is considered to be a basic characteristic of flows in nozzles. We shall assume that in a small element of the nozzle formed by two cross-sections at right angles to the axis, this gradation of stream lines exists. Thus the flow in such a section will be described by a function which in some respects plays the part of a source placed at the vertex of the truncated cone formed by the two sections. If the slope of the nozzle at a particular section with radius r_0 is δ_0, then we assume that the slope of the stream line through the circle of radius r in that section is (see fig. 1)

$$\delta=\left(\frac{r}{r_0}\right)\delta_0. \qquad (1)$$

Fig. 1.



Variation of direction of stream lines.

In practice, the use of diffusers has been confined to those of conical or almost conical shape, and no experimental evidence on the slope of the stream lines in a general divergent flow is available. Gourzhienko (1939) found that in the flow through a cylindrical-conical-cylindrical pipe the " assumption of radiality " was justified. The flows considered in this paper are confined to pipes which have none of the discontinuities in slope of the type mentioned above, and we should expect that the assumption would be even more justified in these cases.

§ 2. NOTATION.

We take the axis of the tube as the z-axis ; the coordinate defining distance perpendicular to the axis is r, and corresponding spherical coordinates are (R, θ). In irrotational flows, the equation satisfied by the potential is

$$\phi_{rr}+\frac{1}{r}\phi_r+\phi_{zz}=\frac{1}{a^2}(\phi_r^2\phi_{rr}+2\phi_r\phi_z\phi_{rz}+\phi_z^2\phi_{zz}) \quad . \quad . \quad . \quad (2.1)$$

or

$$\phi_{\mathrm{RR}}+\frac{2}{\mathrm{R}}\phi_{\mathrm{R}}+\frac{1}{\mathrm{R}^2}\phi_{\theta\theta}+\frac{\cot\theta}{\mathrm{R}^2}\phi_\theta$$
$$=\frac{1}{a^2}\left(\phi_{\mathrm{R}}^2\phi_{\mathrm{RR}}+\frac{2}{\mathrm{R}^2}\phi_{\mathrm{R}}\phi_\theta\phi_{\mathrm{R}\theta}+\frac{\phi_\theta^2}{\mathrm{R}^4}\phi_{\theta\theta}-\frac{\phi_{\mathrm{R}}\phi_\theta^2}{\mathrm{R}^3}\right), \quad . \quad (2.2)$$

where a is the local speed of sound and is a function of velocity if the pressure is a function of density only. Equations (2.1) and (2.2) are forms of the general equation for the potential in the appropriate coordinates (Durand 1934). If an adiabatic law of the form $p=kp^{\gamma}$ applies, then

$$a^2=\frac{dp}{d\rho}=k\gamma\rho^{\gamma-1},$$

and Bernouilli's equation assumes the form

$$a^2=\frac{\gamma-1}{2}(\mathrm{Q}^2-q^2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

where q is the velocity at Mach number q/a, and Q is the speed of flow into vacuum, which under adiabatic conditions, is the greatest speed attainable. Under isothermal conditions, the speed of sound is constant and equal to a_0 ; Bernouilli's equation is then

$$a_0^2 \ln \rho+\tfrac{1}{2}q^2=\text{constant.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.4)$$

In this case there is no limit to the speed.

§ 3. THE FUNDAMENTAL SUPERSONIC FLOW IN A DIVERGENT CONE.

The solution is exact and the flow is irrotational. The gas state is constant on spheres centred at the source, and the velocity increases along radial lines from this point. The formula relating velocity to distance R is given by the equation of continuity which assumes the form

$$\rho\mathrm{R}^2q=\text{constant.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

From Bernouilli's equation (2.3), we immediately obtain

$$(\mathrm{Q}^2-q^2)^{1/\gamma-1}q\mathrm{R}^2=\text{constant.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.2)$$

Let us now suppose that the velocity at a point denoted by R is u (see fig. 2).

It immediately follows from (3.2) that

$$\frac{u_0}{u}\left(\frac{Q^2-u_0^2}{Q^2-u^2}\right)^{1/\gamma-1}=\left(\frac{R}{R_0}\right)^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.3)$$

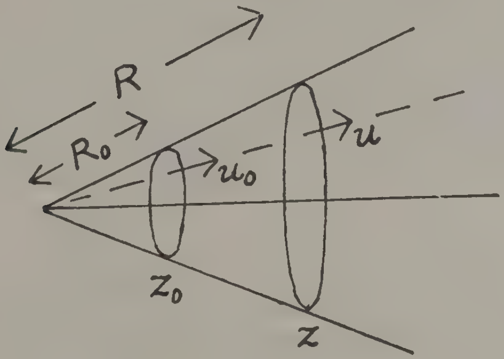
which gives the outlet velocity in terms of the inlet velocity, the speed of flow into vacuum, and the dimensions of the tube.

It should be noted that equation (3.3) holds in both supersonic and subsonic cases; it follows from Bernouilli's equation that in the supersonic case, Q^2 is certainly less than $(\gamma+1/\gamma-1)u_0^2$, and for very high speeds will be considerably less than this. In subsonic flows Q^2 is greater than $(\gamma+1/\gamma-1)u_0^2$.

§ 4. Flow in a Divergent Nozzle.

It is assumed that the equation of the nozzle boundary is $r=f(z)$. Generally, the nozzle will be of the form indicated in fig. 3.

Fig. 2.



Velocity distribution in a conical flow.

The initial velocity on a representative stream line is u_0; the velocity at the point where this stream line intersects a z-constant section is u. It is supposed that the distribution of velocity u_0 across the inlet is known. By use of fig. 1, and a knowledge of u as a function of u_0, we may find the velocity at any point on any streamline. If u_0 is constant, the velocity will be constant across any section perpendicular to the axis, and we may estimate the section at which a particular velocity or a particular Mach number is attained.

We assume that the flow between the two sections z and $z+\delta z$ appear to diverge from the point on the axis with coordinate $z-[f(z)/df/dz]$, or that the slope of the stream lines is proportional to distance from the axis. This assumption in effect determines the equations of the stream lines, since from (1.1) we may write the differential equation of the family as

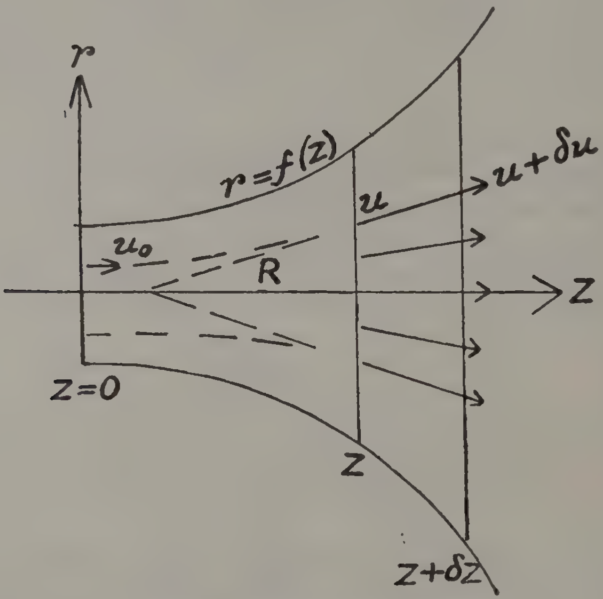
$$\frac{dr}{dz}=\delta=\left(\frac{r}{r_0}\right)\delta_0=\frac{r}{f(z)}\cdot\frac{df}{dz}=r\cdot\frac{d}{dz}(\ln f(z)). \quad . \quad . \quad . \quad (4.1)$$

We immediately obtain the stream lines as the family represented by the equation

$$r=\mu f(z), \quad \ldots \ldots \ldots \quad (4.2)$$

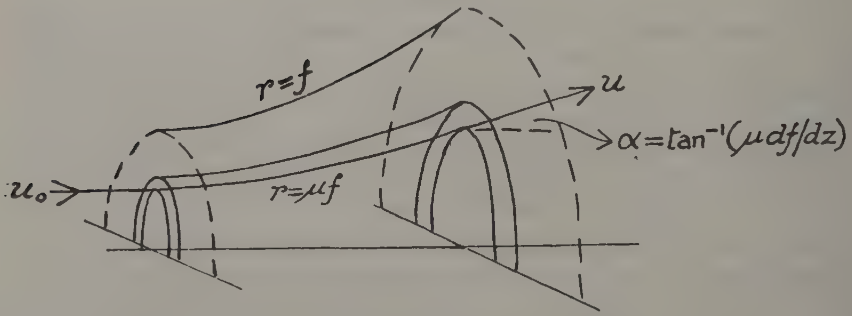
where μ may be any constant between 0 and 1.

Fig. 3.



Velocity distribution in a nozzle.

Fig. 4.



We now consider the motion of the fluid contained in the elementary volume bounded by the stream surfaces $r=\mu f(z)$ and $r=(\mu+d\mu)f(z)$ (see fig. 4). The equation of continuity is now sufficient to determine the velocity u at any point. The flux across the annulet at $z=0$ is $2\pi\{\mu f(0)\}\{d\mu f(0)\}\rho_0 u_0$ which must equal the final value of this quantity, $2\pi\{\mu f(z)\}\{d\mu f(z)\}\rho u\cos\alpha$. We therefore obtain

$$\left(\frac{u_0}{u}\right)\left\{\frac{f(0)}{f(z)}\right\}^2=\frac{\rho}{\rho_0}\cos\alpha=\left(\frac{Q^2-u^2}{Q^2-u_0^2}\right)^{\frac{1}{\gamma-1}}\left\{1+\mu^2\left(\frac{df}{dz}\right)^2\right\}^{-\frac{1}{2}}. \quad . \quad (4.3)$$

In cases of practical interest, the factor $\{1+\mu^2(df/dz)^2\}^{-\frac{1}{2}}$ may be considered to be unity, and the working equation for u assumes the form

$$\left(\frac{u_0}{u}\right)\left(\frac{Q^2-u_0^2}{Q^2-u^2}\right)^{\frac{1}{\gamma-1}}=\left\{\frac{f(z)}{f(0)}\right\}^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.4)$$

As $f(z)\rightarrow\infty$, it is immediately seen that $u\rightarrow Q$. It is noted that the increase in velocity is determined not by the curvature of the nozzle, but only by $f(z)/f(0)$. Thus a short tube with great curvature produces the same increase in velocity as a long tube with small curvature, if the ratios of outlet to inlet radii are equal in the two cases.

It is easy to show that if a divergent flow is being considered $\{[f(z)/f(0)]>1\}$, equation (4.4) has two roots, one lying between $u=0$ and $u=u_0$ and one lying between $u=u_0$ and $u=Q$. In the supersonic case, u is an increasing function of R and the second solution is the relevant one.

Exactly the same procedure may be followed to give the subsonic velocity of a fluid in a divergent flow. Here the velocity ultimately decreases to zero and the root of (4.4) lying between $u=0$ and $u=u_0$ is the relevant one in this case.

§ 5. DIVERGENT FLOWS UNDER ISOTHERMAL CONDITIONS.

When the equation of state is of the form $p=a_0^2\rho$, the local speed of sound is always a_0, and Bernouilli's equation is of the form (2.4). No limiting value of the velocity exists as in the adiabatic case. Proceeding as before, we obtain, in the same notation,

$$\left\{\frac{f(z)}{f(0)}\right\}^2=\left(\frac{u_0}{u}\right)\exp\left\{\frac{1}{2a_0^2}(u^2-u_0^2)\right\}. \quad . \quad . \quad . \quad . \quad (5.1)$$

In terms of Mach number,

$$\left\{\frac{f(z)}{f(0)}\right\}^2=\frac{M_0}{M}\exp\left\{\tfrac{1}{2}(M^2-M_0^2)\right\}. \quad . \quad . \quad . \quad . \quad . \quad (5.2)$$

§ 6. NUMERICAL RESULTS.

For purposes of illustration it is now proposed to calculate the velocity u at the point at which some representative stream line intersects the particular section of the nozzle of radius $f(z)$ in terms of the inlet velocity on that stream line at the section of radius $f(0)$. The form of the streamline may be estimated from the considerations of § 1. If the streamline intersects the inlet section at a distance $sf(0)$ from the axis the equation of the streamline is, according to our assumptions, $r=sf(z)/f(0)$.

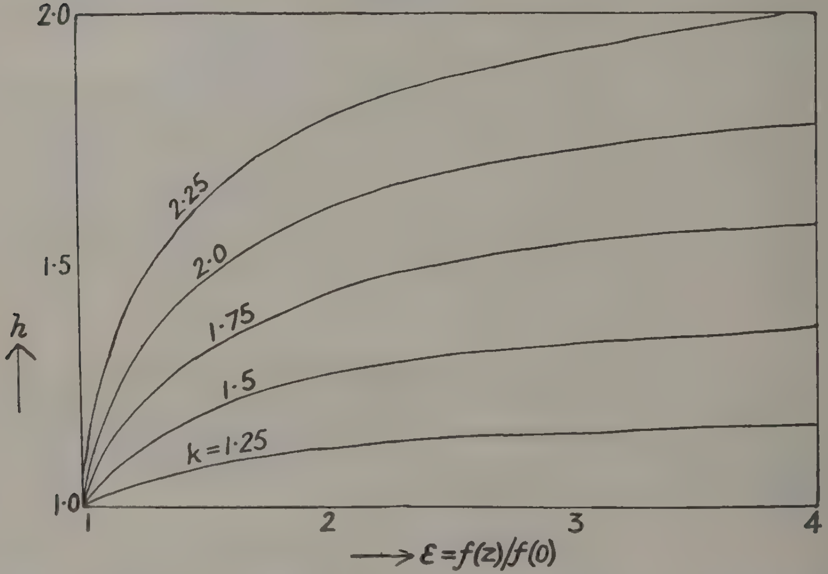
We suppose that the inlet velocity on the chosen streamline is u_0, the velocity on the stream line at the point whose abscissa is z is hu_0, and Q, which is an absolute constant for the whole field of flow, is ku_0. In supersonic flows, k is always less than $(\gamma+1/\gamma-1)^{\frac{1}{2}}$, whilst for subsonic flows, it is always greater than this value. The inlet local Mach number is obtained from Bernouilli's equation and is

$$\left(\frac{2}{\gamma-1}\,\frac{h^2}{k^2-h^2}\right)^{\frac{1}{2}}$$

The numerical work which now arises is the tabulation of a form of equation (4.4)

$$\frac{1}{h}\left(\frac{k^2-1}{k^2-h^2}\right)^{1/\gamma-1}=\left\{\frac{f(z)}{f(0)}\right\}^2=\epsilon^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.1)$$

Fig. 5.



where ϵ may be interpreted as the expansion ratio of the nozzle with reference to section z. If we suppose that the initial velocity u_0 and initial Mach number M_0 are known, then $Q=ku_0$ is calculated as $\{1+[2/(\gamma-1)M_0^2]\}^{\frac{1}{2}}u_0$. We may thus find the value of $f(z)/f(0)$ from equation (6.1) which will produce a velocity on that streamline of $u=hu_0$. When this velocity is known, the local Mach number is calculated from Bernouilli's equation.

The following tables give $\epsilon=f(z)/f(0)$ as a function of k and h, assuming $\gamma=1{\cdot}4$. For purpose of tabulation, it is much simpler to tabulate this quantity than to give a set of solutions of the equation for h when k and ϵ are known. From the results of Table I. a graph has been prepared giving values of $u=hu_0$ as a function of the two variables $Q=ku_0$ and $\epsilon=f(z)/f(0)$ (fig. 5).

The tables are subdivided into four sections involving subsonic and supersonic flows under adiabatic and isothermal conditions. As example of their use :

For simplicity let us assume the velocity to be constant across sections at right angles to the axis. (If this is not so, we have merely to build up the velocity field from considerations of velocities on particular streamlines.) If $Q=1{\cdot}5u_0$, under adiabatic conditions, the initial Mach number (denoted in brackets in Table I.) is 2. In a nozzle of expansion ratio 2·393, a velocity $1{\cdot}3u_0$ is attained at the exit, where the Mach number is 3·886. Table I *a*. refers to a similar flow under isothermal conditions. An inlet velocity u_0 at Mach number 2 is increased to an exit velocity $1{\cdot}3u_0$ in a nozzle of expansion ratio 1·788. The Mach number at the exit is in this case 2·6.

Thus a comparison is made between isothermal and adiabatic flows at the same inlet Mach numbers.

TABLE I.

Tabulation of $f(z)/f(0)=\epsilon$ (supersonic, adiabatic). Mach numbers are denoted in brackets. Results are given to four significant figures.

$k=\frac{Q}{u_0}$ $h=\frac{u}{u_0}$	1·25	1·5	1·75	2·0	2·25
1·0	1·0 (2·981)	1·0 (2·000)	1·0 (1·557)	1·0 (1·291)	1·0 (1·109)
1·1	1·709 (4·141)	1·200 (2·412)	1·090 (1·807)	1·044 (1·472)	1·019 (1·253)
1·2	6·112 (7·652)	1·570 (2·982)	1·232 (2·106)	1·113 (1·677)	1·053 (1·409)
1 3	—	2·393 (3·886)	1 459 (2·480)	1·216 (1·921)	1·107 (1 582)
1·4	—	5·249 (5·813)	1·849 (2·982)	1·369 (2·192)	1·184 (1 777)
1·5	—	∞ (∞)	2·615 (3·719)	1·602 (2·535)	1·289 (2·000)
1·6	—	—	4·613 (5·044)	1·978 (2·982)	1·448 (2·261)
1 7	—	—	17·40 (9·141)	2·659 (3·609)	1 677 (2·579)
1·8	—	—	—	4·145 (4·617)	2·030 (2·982)
1·9	—	—	—	9·292 (6·990)	2·622 (3·621)
2·0	—	—	—	∞ (∞)	3·779 (4·337)
2·1	—	—	—	—	6·781 (5·812)
2·2	—	—	—	—	25·47 (10·42)

TABLE II.

Tabulation of ϵ (subsonic, adiabatic). Mach numbers given to three significant figures.

$k=\frac{Q}{u_0}$ $h=\frac{u}{u_0}$	3·0	3·25	3·5	3·75	4·0
1·0	1·0	1·0	1·0	1·0	1·0
	(0·791)	(0·723)	(0·667)	(0·620)	(0·577)
0·8	1·058	1·068	1·074	1·080	1·085
	(0·629)	(0·568)	(0·525)	(0·488)	(0·456)
0·6	1·173	1·191	1·204	1·215	1·226
	(0·456)	(0·420)	(0·389)	(0·362)	(0·339)
0·4	1·396	1·423	1·445	1·462	1·477
	(0·301)	(0·277)	(0·257)	(0·247)	(0·225)
0·2	1·941	1·985	2·018	2·046	2·069
	(0·149)	(0·132)	(0·128)	(0·119)	(0·112)
0·0	∞	∞	∞	∞	∞
	(0)	(0)	(0)	(0)	(0)

TABLE I *a*.

Tabulation of ϵ (supersonic, isothermal, with same initial Mach numbers as in adiabatic case).

$h=\frac{u}{u_0}$ ↓					
1·0	1·0	1·0	1·0	1·0	1·0
		(2·0)	(1·557)	(1·291)	(1·109)
1·1	1·520	1·176	1·082	1·041	1·017
	(2·981)	(2·2)	(1·713)	(1·420)	(1·220)
1·2	2·427	1·417	1·192	1·096	1·045
	(3·279)	(2·4)	(1·868)	(1·549)	(1·331)
1·3	4·066	1·788	1·332	1·169	1·084
	(3·577)	(2·6)	(2·024)	(1·678)	(1·442)
1·4	7·134	2·208	1·513	1·261	1·136
	(3·875)	(2·8)	(2·180)	(1·807)	(1·553)
1·5	13·12	2·849	1·742	1·374	1·199
	(4·173)	(3·0)	(2·336)	(1·937)	(1·664)
1·6	25·30	3·761	2·031	1·514	1·277
	(4·472)	(3·2)	(2·491)	(2·066)	(1·775)
1·7	51·07	5·077	2·413	1·686	1·373
	(4·770)	(3·4)	(2·647)	(2·195)	(1·885)
1·8	108·1	7·001	2·898	1·895	1·486
	(5·068)	(3·6)	(2·803)	(2·324)	(1·996)
1·9	239·3	9·865	3·530	2·151	1·620
	(5·366)	(3·8)	(2·958)	(2·453)	(2·107)
2·0	554·9	14·20	4·359	2·467	1·780
	(5·664)	(4·0)	(3·114)	(2·582)	(2·218)
2·1	1347·0	20·88	5·455	2·856	1·972
	(5·962)	(4·2)	(3·270)	(2·711)	(2·329)
2·2	3420·0	31·37	6·916	3·338	2·199
	(6·260)	(4·4)	(3·425)	(2·840)	(2·440)

TABLE II *a*.

Tabulation of ϵ (subsonic, isothermal with same initial Mach numbers).

$h=\frac{u}{u_0}$	1·0	1·0 (0·791)	1·0 (0·723)	1·0 (0·667)	1·0 (0·620)	1·0 (0·577)
↓	0·8	1·056 (0·633)	1·066 (0·579)	1·073 (0·533)	1·079 (0·495)	1·084 (0·462)
	0·6	1·168 (0·474)	1·188 (0·434)	1·202 (0·400)	1·214 (0·371)	1·224 (0·346)
	0·4	1·378 (0·316)	1·417 (0·289)	1·440 (0·267)	1·459 (0·248)	1·474 (0·231)
	0·2	1·924 (0·158)	1·972 (0·145)	2·010 (0·133)	2·039 (0·124)	2·064 (0·116)
	0·0	∞ (0)	∞ (0)	∞ (0)	∞ (0)	∞ (0)

§7. PROPAGATION OF CERTAIN VELOCITY PROFILES IN NOZZLES.

A problem of interest to design engineers is the investigation of a non-uniform velocity profile in a divergent nozzle. This problem cannot properly be dealt with by one-dimensional theory as the variation of velocity across any section may be considerable. The solution may be obtained, however, from the Tables I. and II., and fig. 5. A particular problem is given below to illustrate the method.

Let us suppose that at the inlet of the nozzle, radius R_1, we have a velocity distribution with magnitudes $(1, 0·98, 0·95, 0·87, 0·75)u_0$ at the points with radial coordinates $(0, 0·25, 0·5, 0·75, 1)R_1$, where u_0 is some given velocity. In addition we suppose the flow to be supersonic, with a Q-value of $1·5u_0$. The inlet distribution is thus characterized by values of k equal to (1·5, 1·53, 1·58, 1·73, 2) at the quarter radii. For the purposes of the problem, we require the velocity distributions across the sections of radii R_2 and R_3 where the expansion ratio is 1·5 and 2 respectively. These may be found as follows.

At $0·5R_1$ across the inlet, $k=1·58$. By interpolation from fig. 5, we find that if $k=1·58$, and $\epsilon=2$, $h=1·41$. Thus the velocity at $0·5R_3$, where the expansion ratio is 2, is $(0·95u_0\times 1·41)$, or $1·23u_0$. By following this procedure we may estimate the velocities at the quarter radii of the exit sections in terms of u_0. For large expansion ratios, we should expect the velocity distribution to become uniform and equal to $Q=1·5u_0$. The following results are obtained, to two places of decimals :

$\epsilon=1·0$: $(1, 0·98, 0·95, 0·87, 0·75)u_0$ at $(0, 0·25, 0·5, 0·75, 1)R_1$,

$\epsilon=1·5$: $(1.18, 1·18, 1·16, 1·13, 1·10)u_0$ at $(0, 0·25, 0·5, 0·75, 1)R_2$,

$\epsilon=2·0$: $(1·25, 1·25, 1·24, 1·23, 1·20)u_0$ at $(0, 0·25, 0·5, 0·75, 1)R_3$.

§8. COMPARISON WITH ONE-DIMENSIONAL THEORY.

We may first recapitulate the results of one-dimensional approximate theory relating to flows in nozzles. Whatever the shape of the nozzle, one-dimensional theory assumes the velocity to be constant across sections perpendicular to the axis and that this velocity is wholly axial in direction. This immediately affords the result that

$$\{f(z)\}^2\rho q=\text{constant}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.1)$$

or, using Bernouilli's equation,

$$\{f(z)\}^2 q(Q^2-q^2)^{1/\gamma-1}=\text{constant}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.2)$$

a form which is identical to that of (4.4). Thus, when the velocity is constant across sections of constant z, the theories agree with the proviso that the velocity calculated from one-dimensional theory acts at a slope $\delta=(r/r_0)\delta_0$ where δ_0 is the slope of the nozzle at that section.

In addition, however, the theory outlined in this paper affords values for the fluid velocity and Mach number in flows in which the variation of velocity across a section is considerable. Such a case cannot be solved by the one-dimensional approach.

§9. COMPARISON WITH EXPERIMENT.

It is desirable that comparison should be made of the results derived in this paper with the results of practical experiment. Application was made to the National Physical Laboratory and to the Royal Aircraft Establishment, Farnborough, but unfortunately, very little experimental evidence is available for comparison.

It would also be interesting to compare the results given with those obtained from more exact treatment by the method of characteristics and it is hoped to carry out such an investigation at a future date.

ACKNOWLEDGMENTS.

The author is indebted to Professor L. Rosenhead for valuable criticism in the preparation of this paper. During this period the author was in receipt of a grant from the Department of Scientific and Industrial Research.

REFERENCES.

DURAND, W. F., 1934 (Ed.), *Aerodynamic Theory*, **3**, p. 230.

GOURZHIENKO, Z. A., 1939, Report No. 462, Central Aero-Hydrodynamical Institute, Moscow, reprinted in *N.A.C.A.*, T.M. 1137.

XXXIII. *Noise in Current-carrying Ohmic Conductors.*

By BERNARD MELTZER*.

[Received December 19, 1949.]

ABSTRACT.

The mean-square noise current in a conductor which is in temperature equilibrium and carries a current I, is calculated on the assumption that the current is due to elementary carriers of charge e. The value obtained for the spectral density is $\{(2kT/R)+[2kT/R)^2+(2eI)^2]^{\frac{1}{2}}\}$, where R, T, k are respectively resistance, absolute temperature and Boltzmann's constant. The interpretation of this result is that, over and above the ordinary thermal Nyquist noise, there is present an additional noise, of which the spectral density varies in square-law fashion according to $(e^2R/kT)I^2$ for small I, and in linear fashion according to $\{(-2kT/R)+2eI\}$ for large I. These predictions are compared with Bernamont's experimental results for thin metallic conductors, and the discrepancies are discussed.

INTRODUCTION.

IN a recent paper (Meltzer 1949) the ordinary thermal noise in a conductor, sometimes called Johnson or Nyquist noise, was successfully interpreted as pure shot noise. This was done with the aid of an expression, derived there,

$$I_0=(kT/Re), \qquad (1)$$

for the average current carried, in either direction, across any cross-section of a conductor in thermal equilibrium by the randomly moving elementary charge carriers. Here k is Boltzmann's constant, while T and R are the temperature and resistance of the conductor, which has its terminals short-circuited; and it is assumed, for simplicity's sake, that all the carriers have the same effective charge e. (It was shown that the result is not changed if the latter assumption is waived.) Schottky's law for pure shot noise, applied to $2I_0$, the total average random current, immediately gives the correct value for the Nyquist noise.

The analysis was carried out for the case when there is no source of power in the circuit, so that the nett current is zero. Below is presented the generalization of the analysis to the case when the conductor carries a steady current. The result obtained is that in this case, over and above the Nyquist noise, there is present additional noise which, for small steady currents, varies as the square of the latter, and for large values tends asymptotically to a linear law of variation with steady current.

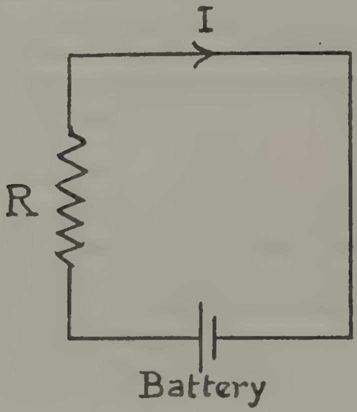
---

* Communicated by the Author.

This appears to be precisely one of the main features of the extra noise observed in thin metallic conductors (Bernamont 1934a, 1934b, 1937), and the square-law variation seems to be characteristic also of the noise observed in semi-conductors, according to Campbell and Chipman (1949). The other chief feature of this additional noise, its variation with frequency, is not predicted by the theory in its present state of development, although a suggestion is given below of how a fuller theory may account for it. A detailed quantitative comparison with Bernamont's experimental results, discussed below, shows that there are some suggestive and encouraging agreements with the theory as to order of magnitude, but by no means perfect overall agreement. The discrepancies are discussed.

ANALYSIS OF FLUCTUATIONS IN A STEADY CURRENT IN A RESISTANCE.

Fig. 1.



Let the steady current$=$I.

If Q is the nett amount of charge carried across any cross-section of the conductor in time t, the mean square value of Q^2 is given by a simple generalization of Einstein's result, quoted in the author's previous paper :

$$\overline{Q^2}=(2kT/R)t+I^2t^2. \qquad (2)$$

Let us analyse Q, assuming that the elementary carriers all carry a charge e. We have

$$Q=e_1+e_2+\ldots+e_i+\ldots+e_n, \qquad (3)$$

where n is the number of elementary carriers that cross the cross-section in time t, and

$$e_i=\pm e,$$

the positive or negative sign depending on whether the particular carrier concerned was going, say, from left to right, or from right to left.

Let $p=$chance that $e_i=+e$, and $q=$chance that $e_i=-e$, so that

$$I=(p-q)(en/t). \qquad (4)$$

Now, from equation (2),

$$Q^2 = \Sigma e_i^2 + \sum_{i \neq j}\Sigma e_i e_j.$$

Hence

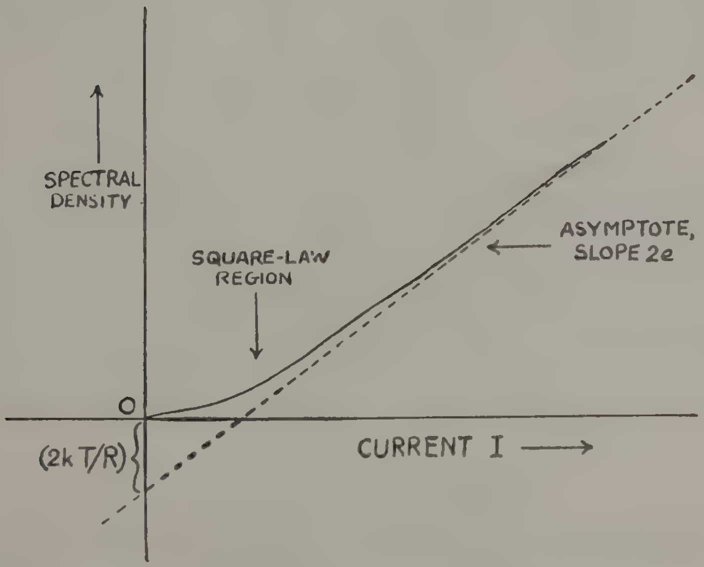
$$\overline{Q^2} = \Sigma \overline{e_i^2} + \sum_{i \neq j}\Sigma \overline{e_i e_j}$$

$$= ne^2 + n(n-1)(p-q)^2 e^2. \quad \ldots \ldots \quad (5)$$

Equating the right-hand sides of equations (2) and (5), gives

$$(ne/t) = (2k\mathrm{T}/\mathrm{R}e) + \mathrm{I}^2(ne/t)^{-1}. \quad \ldots \ldots \quad (6)$$

Fig. 2.



Now (ne/t) is the gross amount of charge carried per unit time across the cross-section by all the carriers, those going from left to right and those going from right to left. Hence to obtain the noise produced in a given frequency band, f to $f+\Delta f$, we must apply Schottky's formula ($2e\Delta f$ times current) to this current. Solving equation (6) gives

$$(ne/t) = (k\mathrm{T}/\mathrm{R}e) + [(k\mathrm{T}/\mathrm{R}e)^2 + \mathrm{I}^2]^{\frac{1}{2}}.$$

But we have seen that the amount of current $(2k\mathrm{T}/\mathrm{R}e)$ accounts for Nyquist noise, hence the additional noise is associated with the current

$$-(k\mathrm{T}/\mathrm{R}e) + [(k\mathrm{T}/\mathrm{R}e)^2 + \mathrm{I}^2]^{\frac{1}{2}}.$$

The application of Schottky's formula gives the mean-square additional noise current

$$\overline{(i_f)^2} = \{-(2k\mathrm{T}/\mathrm{R}) + [(2k\mathrm{T}/\mathrm{R})^2 + (2e\mathrm{I})^2]^{\frac{1}{2}}\}\Delta f. \quad \ldots \quad (7)$$

The graph of spectral density of this noise is shown in fig. 2.

At first, for small I, the spectral density follows a square law given by

$$\overline{(i_f)^2}/\Delta f=(e^2\mathrm{R}/k\mathrm{T})\mathrm{I}^2\,; \quad \ldots \ldots \quad (8)$$

and then, for large I, it approaches a linear, but not proportional, law given by

$$\overline{(i_f)^2}/\Delta f=-(2k\mathrm{T}/\mathrm{R})+2e\mathrm{I}. \quad \ldots \ldots \quad (9)$$

COMPARISON WITH EXPERIMENTAL RESULTS.

Bernamont's first two papers (1934a, 1934b), although historically important, may be ignored for our purposes, since—according to his third paper (1937)—the results are vitiated by a failure to take into account appreciable changes of resistance in the platinum wire used.

In the third paper he dealt with three specimens listed below :

A : Tungsten (" Loewe " resistance) ; 518,000 ohms.
B : Tungsten (" Loewe " resistance) ; 190,000 ohms.
C : Platinum (" platine blanc ") ; 31,200 ohms.

The results we shall consider are for the total additional mean-square noise voltage appearing across the terminals of the specimen over an effective frequency band $\Delta f=1200$ cycles per second, close to the zero end of the frequency scale. The ranges of steady current used and mean-square noise voltage observed for the three specimens are as follows :

A : 0·3 to 137 microamperes ; 28 to $2{\cdot}7\times10^5$ (microvolts)^2.
B : 48 to 490 microamperes ; 9 to 830 (microvolt)^2.
C : 10 to 750 microamperes ; 0·9 to 2700 (microvolt)^2.

If mean-square voltage $\overline{(v_f)^2}$ is plotted against steady current I on linear scales, the curves for all three specimens have the general shape of the graph of fig. 2, except that the one for B does not extend into the approximately linear region. The asymptotic slopes of the other two are given by the following :

A : $2{\cdot}36\times10^{14}$ e.m.u.
C : $4{\cdot}75\times10^{11}$ e.m.u.

Now according to equation (9) of the theory above this slope should be given by

$$2e\mathrm{R}^2\Delta f.$$

Hence we may calculate effective values of the elementary charge e :

A : $e=0{\cdot}37\times10^{-20}$ e.m.u.
C : $e=20\times10^{-20}$ e.m.u.

These are to be compared with the electronic charge of $1 \cdot 6 \times 10^{-20}$ e.m.u.

If the results be re-plotted on logarithmic scales, one can obtain the law of variation over the initial non-linear portions of the curve :

$$\text{A}: \ \overline{(v_f)^2} = 2 \cdot 88 \times 10^{20} \text{I}^2 \text{ e.m.u.}$$
$$\text{B}: \ \overline{(v_f)^2} = 3 \cdot 80 \times 10^{15} \text{I}^2 \text{ e.m.u.}$$
$$\text{C}: \ \overline{(v_f)^2} = 7 \cdot 94 \times 10^{15} \text{I}^2 \text{ e.m.u.}$$

According to equation (8), the law should be

$$\overline{(v_f)^2} = (e^2 \text{R}^3 \Delta f / k\text{T}) \text{I}^2 \, ;$$

hence effective values of e may again be calculated (assuming T=290 degrees Kelvin) :

$$\text{A}: \ e = 0 \cdot 83 \times 10^{-20} \text{ e.m.u.}$$
$$\text{B}: \ e = 0 \cdot 014 \times 10^{-20} \text{ e.m.u.}$$
$$\text{C}: \ e = 0 \cdot 30 \times 10^{-20} \text{ e.m.u.}$$

It is seen that for A and C there are discrepancies respectively of factors of 2 and 67 in the values of elementary charge effective over the square-law and linear regions, respectively, of the observations. On the other hand, all the values, except the single one for B, are within an order of magnitude of the electronic charge. [Another test of the theory which is less encouraging, however, is the following : On the linear scale graphs of A and C, the asymptotes should according to equation (9) cut the current axis at $\text{I} = (k\text{T}/\text{R}e)$, so that another estimate of effective e may be made from the values of the intercepts. This gives :

$$\text{A}: \ e = 3 \cdot 4 \times 10^{-23} \text{ e.m.u.}$$
$$\text{C}: \ e = 6 \cdot 8 \times 10^{-23} \text{ e.m.u.}]$$

It is therefore clear that the analysis given above is inadequate as a complete quantitative explanation of Bernamont's results. Furthermore it does not predict any frequency dependence of the noise, which is an experimentally determined fact. In fact it seems probable that it is just this defect of the theory that is responsible for a large part of the discrepancies calculated above. For, some measurements carried out by Bernamont on the platinum specimen C showed a drop in average spectral density by a factor of 1000 for a 1000-fold increase in frequency ; since the effective pass band of the experiments considered above cover as much as 1200 cycles per second at the lower end of the frequency scale, it is clear that it would have been more appropriate to use results over a much smaller frequency band to check the theory.

A possible explanation of the frequency dependence has been suggested by Dr. R. Fürth in discussion. The effect might be similar to Debye and Falkenhagen's account (1928) of the frequency dependence of the conductivity of electrolytes. For the effective charge of an elementary particle which crosses a given cross-section may be not merely its own charge e but a value modified by the existence of a charged " atmosphere " around it. In electrolytes, such an atmosphere is created because the Coulomb forces exerted, say, by a given positive ion produce in its neighbourhood an excess of negative ions and a defect of positive ones, on the average ; for semi-conductors one might expect something similar, electrons taking the place of negative ions and holes the place of positive ions ; in the case of metals, the main effect is presumably a defect of electrons in the neighbourhood of a given electron. If the relaxation time associated with the production of such an atmosphere during the movement of the particle or hole concerned, is not very much smaller than the period of the frequency, one would expect frequency-sensitive effects. On the face of it, such effects should produce a lowering of effective charge, and it is therefore perhaps significant that of the seven values of charge calculated above, all but one are smaller than the charge of an electron.

The discrepancies described above may also, possibly, be associated with experimental error, or the assumption of only one kind of effective charge. The theory can be re-stated for a number of different types of charge carriers, just as was done for the case $I=0$ in the author's previous paper, but does not then lend itself to comparison with the rather sparse experimental data publicly available at present.

The author wishes to thank Dr. R. Fürth for stimulating and helpful advice.

REFERENCES.

BERNAMONT, J., 1934 a, *Compt. Rend.* (Paris), **198,** 1755 ; 1934 b, *Ibid.*, **198,** 2144 ; 1937, *Ann. de Phys.*, Ser 11, **7,** 71.
CAMPBELL, R. H., and CHIPMAN, R. A., 1949, *Proc. I.R.E.*, **37,** 938.
DEBYE, P., and FALKENHAGEN, H., 1928, *Phys. Zs.*, **29,** 121.
EINSTEIN, A., 1906, *Ann. Phys. Lpz.*, **19,** 371.
MELTZER, B., 1949, *Phil. Mag.*, **40,** 1224.

XXXIV.—*A 4-dimensional Generalization of Wilson's Hypothesis.*

By A. Papapetrou,
I.C.I. Research Fellow, University of Manchester*.

[Received January 31, 1950.]

Summary.

Wilson's hypothesis of electric currents accompanying the rotation of electrically neutral matter (in order to explain the origin of the geomagnetic field) is generalized in a 4-dimensional language. The main consequences of the generalized formulation are discussed.

Several attempts at formulating a unified theory of gravitation and electromagnetism have been made, but none of them has led to satisfactory results. In spite of this failure, it seems to be almost certain that, if a correlation of the two fields really has a physical meaning, it must, first of all, explain the origin of the magnetic fields of the earth, the sun and other stars in the way recently proposed by Blackett (1949). It therefore might be useful to start the discussion from the other end, and try to see whether the demand of an explanation of the geomagnetic field could give some hints about the nature of the dependence of the electromagnetic on the gravitational field. A simple type of correlation of the two fields suggested by such a discussion will briefly be described in the following.

From a phenomenological point of view the geomagnetic field is sufficiently described by Wilson's hypothesis (Wilson 1923). This postulates that a rotating material ring will carry the current density

$$\mathbf{s} = -\frac{1}{c}\sqrt{(\mathrm{G})}\rho\mathbf{v}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

ρ being the mass density, $\mathbf{v}$ the peripheral velocity of rotation, and G the gravitational constant. Hence it will be sufficient to demand from the new postulate, which we are going to formulate, that it must reduce to Wilson's formula (Wilson 1923) when applied to a rotating ring. A second demand which we have to impose is that there must be no contradiction of the general principles of Maxwell's theory. In its simplest form the last demand is equivalent to the condition that the additional "gravitational" sources of the electromagnetic field, which we are going to introduce, must be describable by a charge-current 4-vector density fulfilling the continuity equation

$$\frac{\partial s^i}{\partial x^i} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

* Communicated by Professor L. Rosenfeld.

The main difficulty of Wilson's hypothesis is that it makes use of the *three*-dimensional velocity $\mathbf{v}$. The relativistic point of view demands a 4-dimensional formulation, and a straightforward generalization of Wilson's hypothesis for four dimensions would lead to an additional and extremely strong electric field which does not exist in nature. A simple way out of this difficulty is to correlate any given element of mass with a corresponding *central point*, and then to introduce the *difference* of the 4-dimensional velocities of the mass element and the corresponding central point into the formula defining the gravitational sources of the electromagnetic field. Such a central point is immediately suggested in Newton's theory of gravitation when we consider the gravitational field of the earth. The equipotential surfaces $\phi=$const. are almost spheres and there is in the interior of the earth one equipotential surface which shrinks to a point. We shall call this point the *gravitational centre* of the earth. There is a large number of such centres in space. The moon, the planets, their satellites if they are not too small, the sun, the stars, each of them has its own gravitational centre. Any line of force of the gravitational field starts from *one* of these centres and goes to infinity without meeting any other centre *. Thus an arbitrary mass element is, at a given time, correlated to one gravitational centre; in order to find this centre it is sufficient to follow the gravitational line of force passing through the actual position of the mass element in the direction of the gravitational force acting on the mass element. It is clear that this definition of the gravitational centre will be valid in the 4-dimensional treatment too, provided that there exists a *scalar* function corresponding to the Newtonian potential of gravitation. This possibility exists even in Einstein's theory of gravitation (general relativity) when interpreted as a theory of gravitation in a flat space (Rosen 1940, Papapetrou 1948). Besides g_{ik}, which is then only the potential of gravitation, we have the metrical tensor η_{ik} of flat space, and there is the scalar

$$\phi=\eta^{ik}(g_{ik}-\eta_{ik})$$

which differs very little from the Newtonian potential of gravitation. (There seems to be no possibility of defining the gravitational centre in the usual form of general relativity working with Riemannian space only.)

Having defined the gravitational centre corresponding to a given mass element at a given time, we can now immediately formulate our postulate. We call u^i the 4-dimensional velocity of the mass element:

$$u^i=\left\{\frac{\mathbf{v}}{c\sqrt{(1-\beta^2)}};\ \frac{1}{\sqrt{(1-\beta^2)}}\right\}, \qquad \beta=\frac{v}{c}, \quad . \quad . \quad . \quad (3)$$

---

* Because there are only positive masses. The picture is comparable with that of an electrostatic field arising from charges of one sign only.

and U^i the velocity of the corresponding gravitational centre *. Then we put

$$p^i=\sqrt{(G)}\rho(u^i-U^i), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and introduce the quantity ω fulfilling the equation

$$\Box\omega=\frac{\partial p^i}{\partial x^i}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Finally we define the gravitational charge-current density by the relation

$$s_i=\frac{\partial\omega}{\partial x^i}-p_i, \qquad (p_i=\eta_{ik}p^k). \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

One verifies immediately that the vector s_i thus defined fulfills the continuity equation (2).

It remains to be proved that in the case of the earth our postulate leads to the magnetic field calculated from Wilson's formula (1). We first notice that for all mass elements inside and near the surface of the earth the gravitational centre is a point situated very near to the mass centre (or the geometrical centre) of the earth †. Thus if we use the frame in which the centre of mass is at rest, we shall have

$$U^i=(0, 0, 0, 1).$$

We see that, since $\beta=v/c\ll1$, then

$$u^4=1/\sqrt{(1-\beta^2)}\approx1$$

and consequently according to (4) $p^4\approx0$ and

$$p^i=\{\mathbf{p};\ 0\}, \qquad \mathbf{p}=\frac{1}{c}\sqrt{(G)}\rho\mathbf{v},$$

v being the velocity of the mass element relative the the mass centre of the earth. For a uniform ring around the earth's axis we shall then have

$$\frac{\partial p^i}{\partial x^i}\approx\operatorname{div}\mathbf{p}=0,$$

and consequently from (5)

$$\omega=0.$$

---

* It is easy to see that the position of the gravitational centre will be almost simultaneous with the position of the mass element. The gravitational field of the planets is nearly static (because of their velocities being small compared with c) and consequently the lines of force—*i. e.* the trajectories of the hypersurfaces $\phi=$const—will almost be situated on the hyperplanes $t=$const.

† The main displacement of the gravitational centre of the earth relative to its mass centre is caused by the sun and it amounts to less than 5 km. For the creation of a new gravitational centre near the surface of the earth by an accumulation of mass of the earth's density a body of dimensions not much smaller than those of the earth would be necessary.

It follows from (6)

$$\mathbf{s} = -\frac{1}{c}\sqrt{(G)}\rho\mathbf{v}, \qquad s^4 = 0,$$

i. e. in first approximation we have Wilson's current distribution (1) with no charge distribution present.

From the experimental point of view the case of an isolated body near the surface of the earth will be more interesting than that of a circular ring, since one might hope to be able to detect and measure in the laboratory the magnetic field produced by such a body. We therefore calculate the magnetic field produced by an isolated body according to postulate (6). Since the velocities will be small compared with c, we omit all time derivatives, and find from (5)

$$\Delta\omega = -\operatorname{div}\mathbf{p}.$$

This is the same equation as that giving the potential of a dipole distribution and its solution is

$$\omega = -\frac{1}{4\pi}\int\frac{1}{r^3}(\mathbf{pr})\,dv.$$

We then see from (6) that the lines of current have the geometrical form of the lines of forces of a dipole distribution whose density is proportional to $\mathbf{p}$. From the current density $\mathbf{s}$ we have to calculate the magnetic potential $\mathbf{A}$ by means of the equation

$$\Delta\mathbf{A} = -4\pi\mathbf{s},$$

and then the magnetic field $\mathbf{H} = \operatorname{curl}\mathbf{A}$. Therefore

$$\Delta\mathbf{H} = -4\pi\operatorname{curl}\mathbf{s} = 4\pi\operatorname{curl}\mathbf{p}.$$

The integration of this equation gives

$$\mathbf{H} = -\int\frac{1}{r}\operatorname{curl}\mathbf{p}\,dv,$$

which after a partial integration leads to the formula

$$\mathbf{H} = -\int\frac{1}{r^3}[\mathbf{pr}]\,dv. \quad \ldots\ldots\ldots \quad (7)$$

We see that the magnetic field produced by the mass element contained in the volume element dv is identical with that given by Biot–Savart's formula for a current element $-\mathbf{p}\,dv$.

Since $\mathbf{p}$ is linear in $\mathbf{v}$, it follows that if the body performs a complicated motion which can be reduced to a number of simpler ones, the total magnetic field produced by the body will be the sum of the magnetic fields corresponding to these simpler motions*. The following two examples give characteristic consequences of our postulate (6). A projectile moving relative to its terrestrial surrounding with the velocity **V**, while the peripheral velocity of rotation of the earth at the place of the experiment is $\mathbf{v}_0$, will produce the magnetic field corresponding to the velocity $\mathbf{v}=\mathbf{V}+\mathbf{v}_0$. (If, *e. g.* $V=3v_0$ and both V znd v_0 have the same direction, the moving projectile will produce a field four times stronger than if it were at rest.) A rotor rotating in the laboratory will produce a magnetic field which is the sum of the following two fields: a Biot–Savart field corresponding to the translation of the rotor's mass with the velocity $\mathbf{v}_0$, and a dipole field corresponding to the rotation of the rotor around its axis (the magnetic moment of this dipole being given by Blackett's formula). Since in any experiment we cannot obtain velocities much larger than v_0, we see that the magnetic field will always be very small. These consequences are essentially different from those following from the generalization of Wilson's hypothesis suggested by Blackett (1949a): According to the latter there will be no general superposition of the magnetic fields corresponding to different motions of a body, and there is the possibility of the production of much stronger fields either by suitable motions or by suitable arrangement of other masses around the test body.

---

* Strictly speaking the motion of a body in the laboratory will influence the position of the gravitational centre of the earth, *i. e.* the values of U^i. But this will be vanishingly small because of the smallness of the mass of the test body relative to the total mass of the earth.

More complicated consequences follow from the postulate (6) for certain astronomical cases in which large displacements of the gravitational centre occur. For a qualitative discussion it is sufficient to consider a planet of radius R and a satellite of radius r at the distance a from the planet. Assuming for simplicity that both bodies have the same density, we find by an elementary calculation that the satellite will have its own gravitational centre *only if the distance a is larger than a critical value a_0*, given by the relation

$$(a_0-r)^2=\frac{R^3}{r}.$$

Thus a satellite, moving on an elliptic orbit such that the distance a takes values larger *and* smaller than a_0, will have its own gravitational centre only on the outer part of its orbit (*i. e.* the part for which $a>a_0$). Hence on the outer part of the orbit the satellite will produce the magnetic field corresponding to its own rotation (around its own gravitational centre). On the inner part ($a<a_0$) the gravitational centre of the satellite will coincide with that of the planet; therefore the satellite will also produce the magnetic field corresponding to its translational velocity round the planet*.

Similar considerations will hold for a double star, though in this case further complications will arise if the masses of the two stars are approximately equal (and if the two stars are not fully separated).

---

* This again contrasts with the corresponding consequence of the generalization of Wilson's hypothesis suggested by Blackett (1949a), according to which a body moving freely in a gravitational field will only produce the magnetic field corresponding to its own rotation.

Anyway, the general qualitative conclusion from the postulate (6) is that *large periodic variations of the magnetic field of a double star are to be expected, if the orbits are strongly elliptic.* In this connection it is interesting to recall that Babcock (1948) has observed a star whose magnetic field is periodically changing. But very little is known at present about the internal structure of this star, so that it is difficult to see how this observation could give an argument for or against the postulate (6).

As a final remark we ought to stress that the postulate formulated above does not constitute a unified field theory. It might give a successful description of the one physical feature of a unified theory, namely the production of an electromagnetic field by the gravitational field of moving (neutral) bodies. Further we must stress that the formulation is not purely differential. It involves the position of the gravitational centre of the earth (or generally of the gravitational field surrounding the test body), which can only be determined by solving the equations of the gravitational field. This is a most unusual feature, but we do not think that there are *a priori* arguments against it. It is very difficult to see how one would have to proceed in order to set up a complete theory using the postulate (6) ; but some way might be found, if only there were some direct experimental evidence in favour of the postulate.

REFERENCES.

BABCOCK, H. W., 1948, *Phys. Rev.*, **74**, 489.

BLACKETT, P. M. S., 1947, *Nature*, **159**, 658 ; 1949, *Phil. Mag.*, **40**, 125 ; 1949 a, Discussion at R. Astr. Soc. reported in Observatory, **69**, 100.

ROSEN, N., 1940, *Phys. Rev.*, **57**, 147.

PAPAPETROU, A., 1948, *Proc. Roy. I.* A, **52**, 11.

WILSON, H. A., 1923, *Proc. Roy. Soc.* A, **104**, 451.

XXXV. CORRESPONDENCE

Further Evidence for the Existence of τ-Mesons.

By J. B. Harding,
Imperial College, London *.

[Received February 15, 1950.]

We have recently observed an event in Ilford G5 200 μ "Nuclear Research" Emulsions, and this we have to interpret in terms of the spontaneous decay of a τ-meson, *i. e.* a particle of mass of approximately 1000 m (where m is the electronic mass).

A facsimile drawing of the event is shown. The star centre is 6 mm. from the edge of the plate. The angle from the upward vertical to track K, which has a length of 740 μ from the glass to the star centre, is approximately 85° at the emulsion-glass interface. In fig. 2 we have plotted the grain density (number of grains per 70 μ) v. range from the star centre of tracks a, b, and K, and show for comparison the variation of the mean grain density with *residual range* of thirty protons and ten μ-mesons. Unfortunately, track c passes out of the emulsion/air surface after 30 μ and has a grain density of $29 \pm 8{\cdot}3$ (the error being the standard deviation on the number of grains counted).

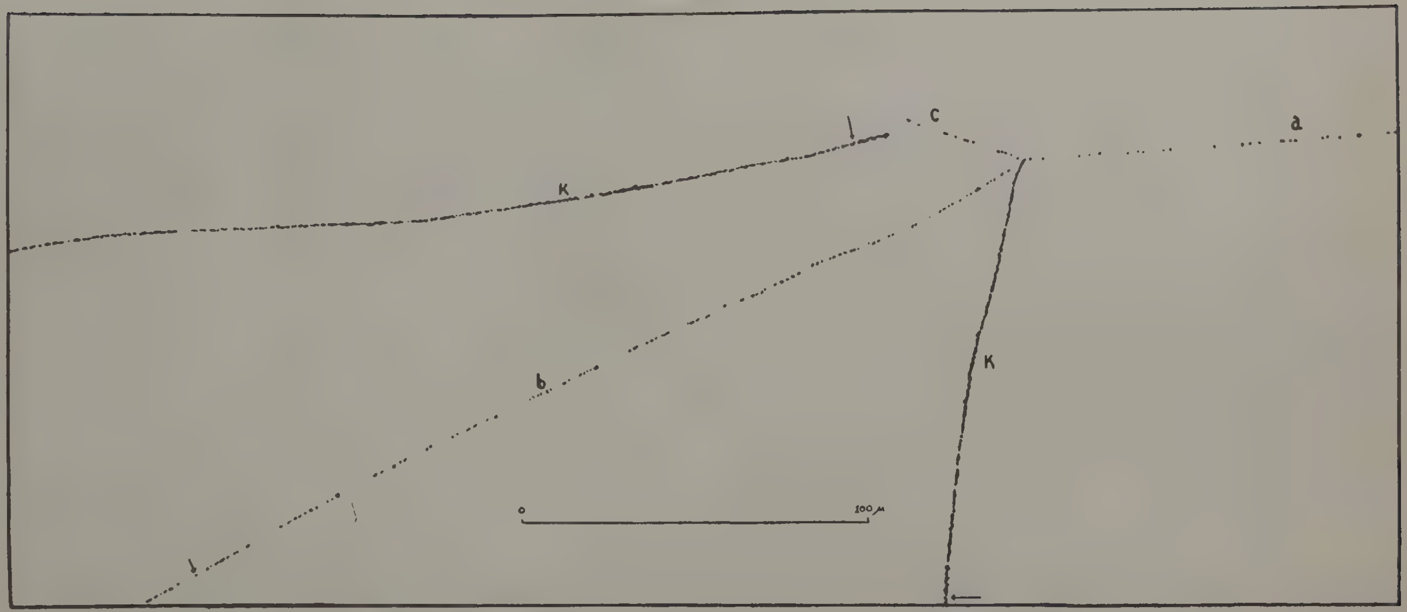
Track b, although it showed considerable multiple scattering, remained almost in the plane of the emulsion for 1300 μ before undergoing a large angle single scatter. It was therefore possible to determine the product momentum × velocity ($\rho\beta$) of the particle which produced this track, by measuring this multiple scattering. The mean projected angle of the track was measured at intervals of 60 μ, and the difference in angle, α, between neighbouring cells thus obtained. The mean of twenty readings of α gave $\bar{\alpha}=1{\cdot}0° \pm 0{\cdot}15°$ and hence we are 99 per cent confident that $16 < \rho\beta < 46$ MeV., the most probable value being 25 MeV. The scattering of track a (range from star to glass $=420\ \mu$) was too small for measurement, and we are thus confident that the spurious scattering is much less than 1/10° per 60 μ. The frequent multiple scattering of track K which increased towards the star centre was more characteristic of a typical π-meson than of a proton.

After allowing for the shrinkage of the emulsion during processing (for details see Harding, Lattimore, and Perkins 1949) it was found that the normal to a plane defined by any two of the tracks, a, b, c, was perpendicular to the third, within 2°, *i. e.* these three tracks were coplanar. Track K did not lie in the same plane.

---

* Communicated by Sir George Thomson.

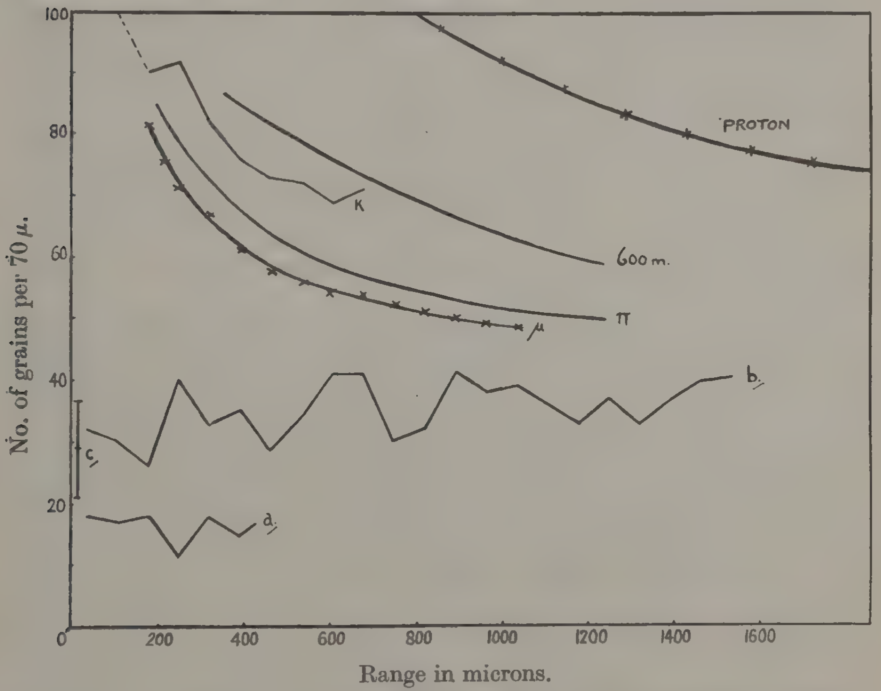
Fig. 1.



Facsimile drawing made with the aid of a projection microscope.
The two sections of track K meet at the arrows.

Before interpreting these experimental results we must first discuss the recording characteristics of the emulsion which was exposed for 85 days under three mètres of ice. In view of the long exposure, there is a certain amount of fading, which, coupled with the low development, makes the recording of tracks at minimum ionization uncertain. We do not feel justified in comparing grain densities of tracks formed in unrelated events, to obtain quantitative information about any particular track. Rather, we must analyse this event on its own merits, and merely include data from other events in order to show that the fading problem cannot alter our conclusions. We have obtained grain density measurements on proton and μ-meson tracks that stop in the emulsion and obtain no significant difference in the grain-densities of those coming from the glass and those from the surface. Therefore the development (temperature method) of the plate is uniform.

Fig. 2.



Both the scattering and grain density of track K decrease with increasing distance from the star centre. We thus conclude that particle K is travelling *towards* the centre of the star and presumably produces it. The grain density measurements indicate that particle K must stop, or at least come very nearly to rest before producing the star. The only generally accepted particles which so behave are π-mesons which give up their rest mass to form stars with excitiation energies of less than 150 MeV. The expected grain density-range curve for π-mesons (as constructed from the curve for

μ-mesons) is also indicated in fig. 2. If, therefore, particle K is a π-meson, this event must have occurred towards the end of the exposure, for the fading must be extremely low as the grain-density of this track is markedly above the average grain density that we would expect for π-mesons. Assuming tracks a, b and c to be formed by protons, we find energies of >300, 110–140, 100–300 MeV. respectively. Thus the total minimum energy of 510 MeV. could not be supplied by the rest mass of a π-meson. We cannot exclude the possibility that track a was formed by a fast electron, but even then its energy must have been greater than 50 MeV. as the scattering was too small to measure. Allowing an improbable fluctuation in grain density, track c could possibly be produced by a slow energy electron. In this way we could obtain a minimum excitation energy of 160 MeV. As yet, no evidence for the production of fast electrons following the nuclear capture of a π-meson has been reported. Also it seems improbable that track b could be formed by a proton of 110 MeV. which would have $\rho\beta=175$ MeV. ($\bar{\alpha}=0{\cdot}15°$), the observed value is only 25 MeV. This track cannot be produced by an electron of 25 MeV. for it should then have minimum grain density, which must be $\leqslant 16$, since we have observed a track of such a grain density in this event.

It seems probable that a meson must have produced track b. There is clearly insufficient energy available for a π-meson to produce a μ-(216 m.) meson (for this is the only lighter one known to exist) and to supply kinetic energy to particle a. In no way then can this event be interpreted along orthodox lines of the nuclear capture or of the spontaneous decay of a π-meson.

The fact that the tracks a, b and c are coplanar is very strong evidence for the spontaneous decay of a particle at rest. Such a process, namely the decay of a τ-meson, of mass $\sim 1000\,m$., into a π^- and probably two other π-mesons, has already been published by the Bristol group (Brown *et al.* 1949). If we assume K represents the track of a τ-meson, the grain density measurements would now show that the latent image fading is by no means negligible and hence energies that we infer for particles a, b and c (assumed now to be π-mesons) cannot be very accurate. It would seem, however, that their energies should lie in the intervals >50, 10–20, 5–10 MeV. respectively. For $\rho\beta=25$ MeV., a π-meson has kinetic energy 13 MeV. Thus for particle b, there is good agreement between the energies estimated from grain counting and scattering measurements. Resolving moments first perpendicular and then parallel to track a, we find that the energies of a and c are 50 and 22 MeV. respectively. Allowing 450 MeV. for the rest energies of the three π-mesons, we find the rest energy of the τ-meson to be 535 MeV. ($\equiv 1040\,m$.).

If we assume the decay products to be μ-mesons, particle b will again have 13 MeV. kinetic energy (for $\rho\beta=25$ MeV.) and the energies of a and c are 49 and 21 MeV. respectively. These are still in agreement with the energies (now >38, 7–15, 4–30 MeV. for a, b and c) that we infer from grain counting. Allowing 330 MeV. for the rest energy of the three mesons, the τ-meson would have a rest energy of 413 MeV. ($\equiv 810\,m$.).

It is seen, therefore, that there is good internal consistency if we interpret this event as the decay at rest of a τ-meson of mass about 800 m. or 1000 m., depending on the decay scheme. Since this event is explained in exactly the same way as was the event of Brown *et al.* (1949), the existence of τ-mesons would seem to be truly confirmed and as such should be duly considered in any meson theory of nuclear forces. From the theoretical point of view, the range of nuclear forces might well correspond to an admixture of π- and τ-mesons. Since the particles are so rarely observed (see for example, Franzinetti 1950, in which negative evidence for the existence of τ-mesons was obtained), they must be rarely produced or have a short lifetime. By now, of the order of 10,000 lightly ionizing particles produced in nuclear disintegration must have been observed in photographic emulsions and yet no evidence for the decay in flight of a τ-meson has been obtained. Assuming that τ-mesons are produced with kinetic energies about three times their rest mass, we conclude that their proper lifetime cannot be much smaller than 10^{-9} secs. if they are to account for a large fraction of these fast particles. With such a lifetime we would then expect to observe more coming to rest and it therefore seems that the hypothesis that they are only rarely produced is necessary for agreement with experimental facts.

I wish to express my appreciation to Professor Sir George Thomson for a helpful and interesting discussion on this event.

Note added in proof.—One further example of a particle coming to rest in the emulsion and producing three nearly coplaner lightly ionising particles, has just been observed. If the incoming particle were assumed to be a π-meson, the fading must be negligible and the three ejected particles, if protons, would have a total kinetic energy of >600 MeV. Two of the tracks could be formed by fast electrons and only in this case would the energy be low enough to be supplied by the rest mass of a π^--meson. The event can easily be explained as the spontaneous decay of a τ-meson into three lighter mesons with total kinetic energy >70 MeV. or >95 MeV. if the decay particles are assumed to be μ- or π-mesons respectively. Slight distortion, and the shortness of the tracks, makes a detailed study of this event impossible.

Department of Physics,
Imperial College, London.

J. B. Harding.

References.

Brown, Camerini, Fowler, Muirhead, Powell, and Ritson, 1949, *Nature*, **163**, 82.
Franzinetti, 1950, *Phil. Mag.*, **41**, 86.
Harding, *et al.*, 1949, *Proc. Roy. Soc.* A, **196**, 325.

Intrinsic Block Structure and the Strength of Metals.

By M. S. PATERSON, Ph.D.,
Aeronautical Research Laboratories, Melbourne*.

[Received February 6, 1950.]

IN his recent paper "On the Strength of Quasi-Isotropic Solids", Dr. Fürth (1949) discusses the idea of an "intrinsic block structure" of vital importance in the processes of melting and rupture. He also suggests that this block structure may be identical with that assumed in Bragg's theory of the strength of metals and the explanation by Wood and co-workers of the X-ray line broadening in cold-worked metals. However, from the proposed form of the dependence of the block size on temperature (increasing with decreasing temperature), Fürth's hypothesis would predict a decrease of X-ray line broadening and strength with decrease of temperature. This is contrary to experiment, for it has been observed that the X-ray line broadening and the stress–strain curve for metals are higher at lower temperatures (Wilson and Thomassen 1934, Hutchison 1948, Paterson and Orowan 1948, Paterson 1950). It is therefore doubtful whether the mechanical properties of a metal are influenced by an intrinsic block structure such as that proposed by Fürth.

It should also be made clear that Bragg's theory deals with the resistance of a metal to plastic flow, whereas Fürth's deals with its rupture. In the case of specimens that "neck" during a tensile test, the "ultimate tensile strength" is a measure of the resistance to plastic flow at a certain point on the stress–strain curve (see, for example, Orowan 1949). Rupture occurs at a later stage in the test, under stresses that cannot be determined from the experiment owing to the complex stress state then existing in the necked region of the specimen.

REFERENCES.

FÜRTH, R., 1949, *Phil. Mag.*, [7], **40**, 1227.
HUTCHISON, T. S., 1948, *Nature*, **162**, 374.
OROWAN, E., 1949, *Rep. Prog. Phys.* (Phys. Soc.), **12**, 185.
PATERSON, M. S., and OROWAN, E., 1948, *Nature*, **162**, 991.
PATERSON, M. S., 1950, to be published.
WILSON, J. E., and THOMASSEN, L., 1934, *Trans. Amer. Soc. Metals*, **22**, 769.

---

* Communicated by the Author.

XXXVI. *Notices of New Books and Periodicals received.*

Theoretical Hydrodynamics. 2nd Edition. By L. M. MILNE-THOMSON. (Macmillan & Co. [600 pp.] Price £3.

THE first edition of this work was published in 1938; in producing the second edition the author has maintained the vector presentation of the subject. For the most part, the book is concerned with the theory of incompressible inviscid flow; there is, however, a chapter on viscosity incorporating a section on boundary layers, as in the former edition, and in addition in the second edition a chapter on compressible flow, subsonic and supersonic.

The presentation of much of the two-dimensional irrotational flow has been modified in the second edition by the use of the "circle-theorem" which though primarily concerned with the effects of introducing a circular cylinder in a given flow has, through conformal transformation, repercussions on aerofoil problems. The corresponding "sphere-theorem" in three dimensional flow is also introduced. The treatment, in the reviewer's opinion, benefits considerably by way of systematization by these changes.

For many problems of inviscid flow vector notation can be used with advantage as here, but particularly in such topics as rate-of-strain and stress in a viscous fluid its limitations are, in the reviewer's opinion, only too obvious.

L. H.

Photoelasticity. By H. T. JESSOP and F. C. HARRIS. [Pp. 184.] (London: Cleaver-Hume Press). Price 28*s*.

THIS is a timely volume on the principles and methods of photoelasticity suitable for the engineer. Since it is written essentially for the experimenter the reader only requires an elementary knowledge of mathematics and no previous knowledge at all of optical polarization on which adequate chapters are included. There are chapters on reduction and interpretation of results and on materials and procedure. The book ends with examples of the use of the method and with some appendices of a rather more mathematical nature.

The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications. By NORBERT WIENER. [Pp. 163]. (New York: Technology Press of M.I.T. and John Wiley & Sons, Inc., 1949). Price $4·00

THIS important and long-awaited monograph is "an attempt to unite the theory and practice of two vitally important fields of work—time series in statistics, and communication engineering". It mainly consists of a basic mathematical theory of linear prediction, linear wave filters, and related problems like approximate differentiation in the presence of "noise", developed by Wiener during the war and first published in a restricted report. The present publication includes also a useful introductory account of the theoretical concepts required, and as appendices two reprinted articles by N. Levinson giving a more elementary mathematical treatment, and a numerical table of Laguerre functions for use with an orthogonal series solution of the prediction problem.

Engineers may find Wiener's discussion rather too abstract for immediate application, though approximations directed towards the representations of ideal solutions by realizable electrical networks are indicated. Mathematicians and statisticians will welcome a much more comprehensive account than that available in the same author's *Cybernetics*, even if the development is in places still somewhat elliptical; one problem, at least in many statistical fields of application, that of first determining precisely the required correlation functions or associated spectra, is not seriously discussed.

M. S. B.

The Physics of Rubber Elasticity. By L. R. G. Treloar. [Pp. vii+255, with 113 text figures.] (Oxford: Clarendon Press 1949.) Price 21*s*.

This is one of the series of Monographs on the Physics and Chemistry of Materials published by the Oxford University Press. It is designed to give an account of recent experimental and theoretical work on the elastic properties of rubber, especially the large elastic extensibility of natural and synthetic rubbers and the type of molecular structure with which it is associated.

The book opens with an account of the general physical properties of rubber and the changes in the thermodynamical functions when a sample of rubber is deformed. This thermodynamical analysis provides evidence of the reversible nature of the deformation of rubbers, and it gives a firm foundation for the statistical developments which follow. These are expounded in the succeeding chapters which deal with the elasticity of long chain molecules and of molecular networks. Different types of strain are then examined, and it is shown that there is good agreement between the experimental results and the statistical theory in the region of small strain, but deviations in the region of very large strain. This topic is considered further in the closing chapter in which an account is given of some of Rivlin's work in this field. There are also chapters on photo-elastic properties, crystallization and mechanical properties, and relaxation phenomena .

The author's remarks following his equation (2·1) are open to objection. He claims that this equation is a definition of internal energy : this claim implies acceptance of the atavistic notion of " heat " which is made untenable by the experiments on which the first law of thermodynamics is based. The experimental basis of the first law is provided by the experiments which Joule described in his memoir of 1849 to the Royal Society. These experiments were done under *adiabatic* conditions. They indicate the *existence* of a function of state which is called the internal energy and the changes in which are equal to the work done on the system under adiabatic conditions. The author's equation (2·1), which refers to *non-adiabatic* processes, can then be used to define " quantity of heat ". It is also a pity that the author has not followed the recommendations of the International Union of Chemistry (1947) and of Physics (1948) in his notation for the thermodynamical functions.

These blemishes do not, however, detract from the very readable account of the properties of natural and synthetic rubbers given in this monograph.

A. R. Miller.

Twelve Lectures on Theoretical Rheology. By Markus Reiner, [Pp. 163.] (North Holland Publishing Company.) Price 9 Dutch florins.

Professor Reiner's well-known " Ten Lectures " have been revised and expanded into the present twelve. The new edition is much more clearly and accurately printed than the old. The main change in content is the adoption of a rigorous tensor notation for the components of strain, with the use of the Kronecker delta to separate the isotropic and deviator components, leading to some simplifications and generalization of the formulæ. The book is an excellent brief account of the mathematical basis of rheology. Further discussions of the connection between " solid " and " liquid " viscosities and of the dimensional basis of " consistency variables " would have been useful, and the argument at the top of page 133 seems to me unconvincing.

F. R. N. Nabarro.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XXXVII. *Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.*—Part IV. *The Distribution in Energy, and the Secondary Interactions of the Particles Emitted from Stars.*

By U. Camerini, P. H. Fowler, W. O. Lock and H. Muirhead.
The H. H. Wills Physical Laboratory, University of Bristol *.

[Received March 20, 1950.]

[Plates IX.–XI.]

SUMMARY.

The measurements of "multiple scattering" and "grain density" discussed in Parts II. and III. of the present series of papers, have now been extended to include a total of 1000 tracks of particles associated with "stars". The method of measurement and the criteria for the selection of tracks were similar to those described previously (III.). The new material provides additional evidence in support of the view that most of the mesons ejected from nuclear explosions with kinetic energy less than 150 MeV. are π-particles.

The distribution in energy of the ejected mesons, protons, deuterons and tritons has been determined, and it is shown that the "spectrum" for the mesons is in good agreement with that calculated by Sands (1950). The mean free-path for nuclear interaction of the various types of particles in passing through the emulsion has been determined. The value obtained for mesons corresponds to a cross-section for the interaction equal to the "geometrical" value.

SCATTERING MEASUREMENTS.

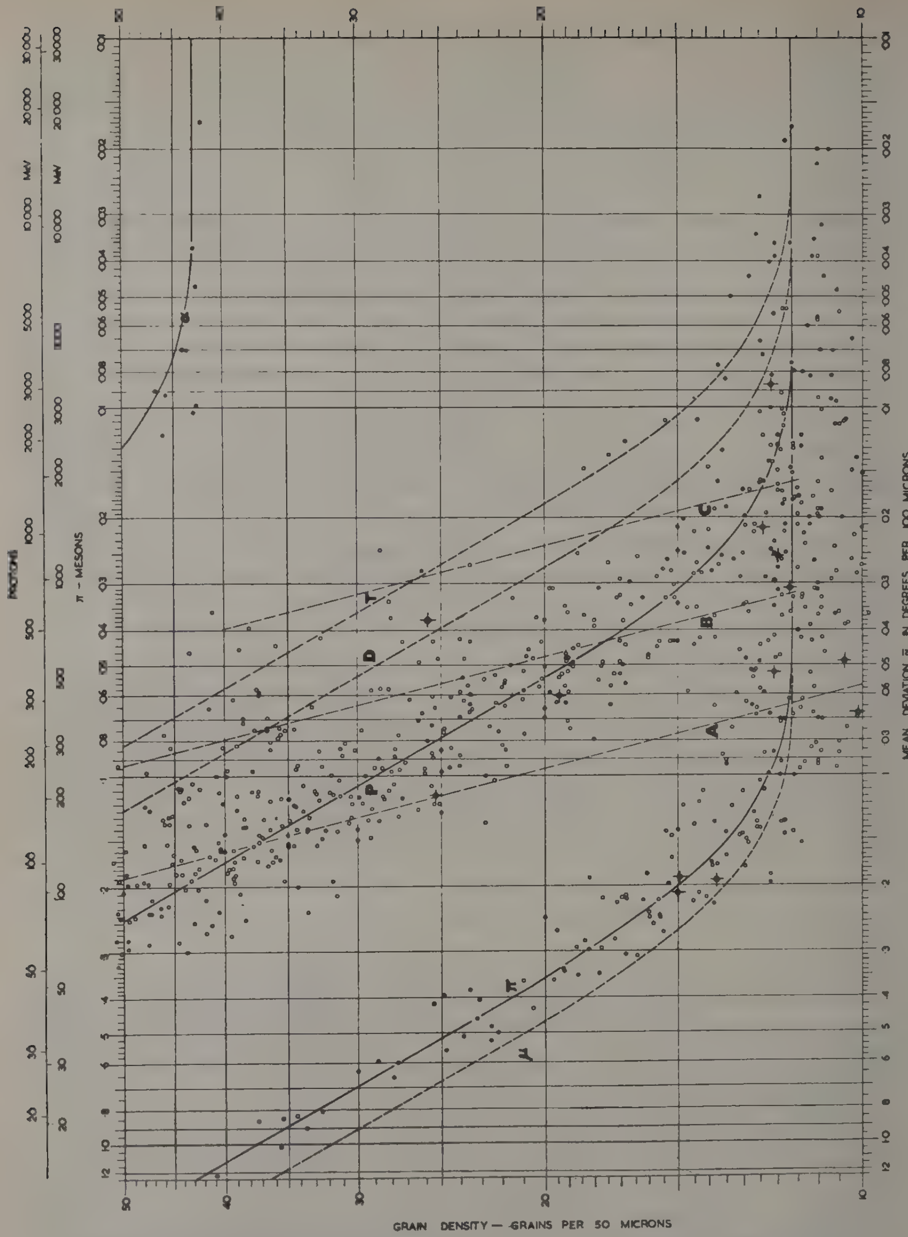
The new measurements were made on plates coated with Ilford G5 emulsion, $400\,\mu$ thick and exposed in the balloon flight represented in curve *c* of fig. 2. Part II. Plates from the same batch were employed in the measurements described in Part III. (Fowler 1950).

The experimental results for 700 tracks, all with a grain density less than 50 grains per $50\,\mu$, are shown in fig. 1. The curves, μ, π, P, D, T are the same as in the corresponding figure in Part III., and the value of the constant k_2 used (equation (3), Part III.) was 32·7. This particular value of k has been chosen to give the best fit with the experimental distribution for those protons which produce tracks with values of the grain-density between 16 and 50 grains per 50μ. For higher values of the grain-density, there is a small departure of the experimental points for protons corresponding to a change in the value of k of about 7 percent. This, however,

---

* Communicated by Professor C. F. Powell, F.R.S.

Fig. 1.



Relation between grain-density and $\bar{\alpha}^\circ$ (100 μ) for the tracks of 700 particles. “ Primary ” particles are shown thus • ; “ Secondary ” particles thus ○ ; and “ interconnecting ” particles thus -ϕ-.

The lines μ . π . P, D, T are the same as in fig. 3 of Part III.

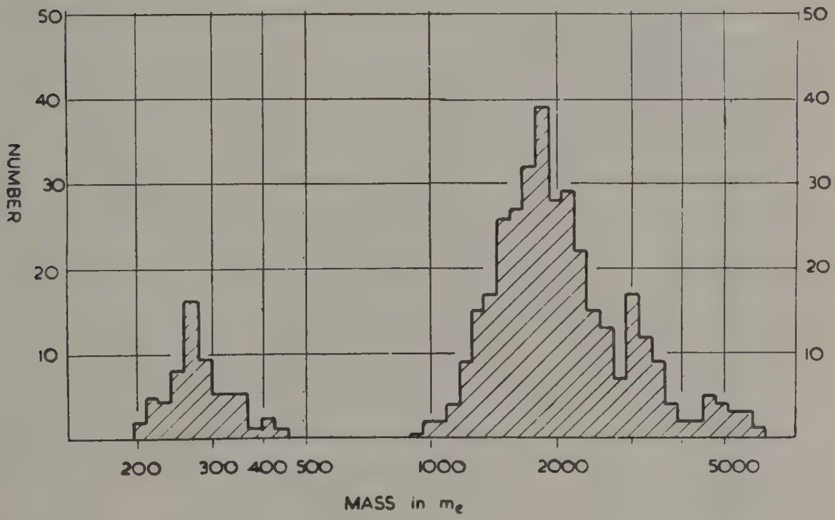
A, B and C are lines of constant momentum and correspond respectively to 500, 1000 and 2000 MeV./c.

Kinetic energy scales for π-mesons and protons are included at the top of the figure.

does not affect the measurements of the mass of the mesons, because a π-particle with a residual range of 3 mm. produces a track with a grain-density of approximately 45 grains per 50μ.

The mass spectrum of the particles has been obtained by the same method as that described in Part III., and is shown in fig. 2. The number of mesons suitable for measurement, and producing tracks with a grain density greater than 15 grains per 50μ is 58, and the average value of the mass of these particles $283 \pm 7\ m_e$.

Fig. 2.



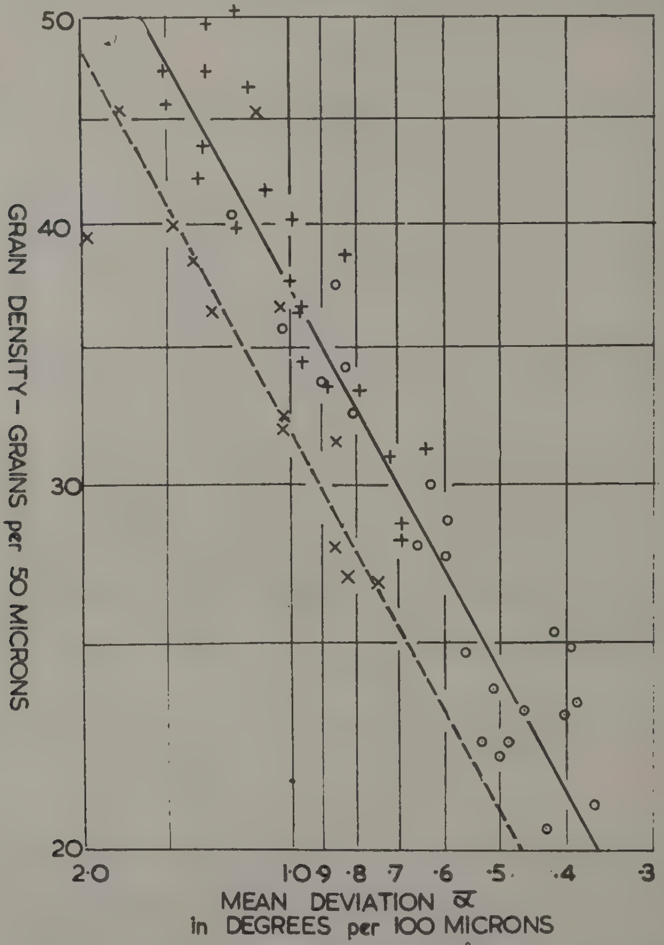
Distribution in mass of the particles of charge /e/ corresponding to the measurements represented in fig. 1.

In order to confirm the identity of these particles, we have made similar meaurements on the tracks of 22 π-mesons and 13 μ-mesons of which the range in the emulsion was greater than 3 mm. The π-mesons were identified by their decay into a μ-meson, or by the occurrence of a nuclear interaction at the end of the range ; and the μ-mesons by the decay into an electron. The results of these measurements are displayed in fig. 3. We believe that these observations leave little doubt that all, or almost all, the mesons ejected from stars with kinetic energies less than 150 MeV. are indeed π-mesons.

The nature of particles with a kinetic energy between 150 MeV. and 1500 MeV., which produce tracks with values of the grain density less than 16 grains per 50μ, cannot be determined from these measurements

The work of Piccioni (1950), however, has shown that most of them are also π-mesons. Further, the possibility of the presence of a large proportion of electrons among the shower particles can be eliminated by the following observation :—Twenty of the 300 shower particles measured have had path lengths in the emulsion greater than 2 cm. Since the radiation length in the emulsion is 2·3 cm. (Carlson *et al.* 1950), a significant decrease in energy should have been observed in at least some of these tracks, had a considerable number of them been due to electrons.

Fig. 3.

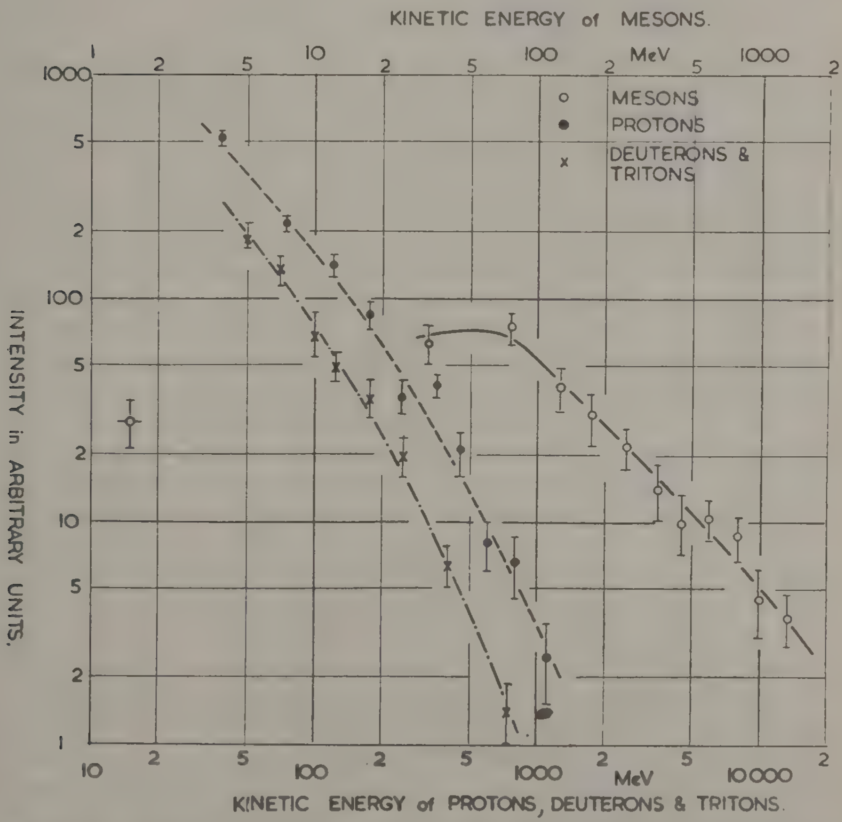


Relation between grain-density and $\bar{\alpha}^\circ$ (per 100 μ). Mesons ejected from stars are shown thus ○, and are the same as those shown in the corresponding grain-density interval in fig. 1. Mesons, which, at the end of their ranges produce stars or π–μ decays are shown thus + ; and mesons which decay into fast electrons at the end of their ranges are shown thus ×.

THE DISTRIBUTION IN ENERGY OF THE PARTICLES EMITTED IN NUCLEAR DISINTEGRATIONS.

The results represented in fig. 1 allow us to determine the distribution in energy of the secondary particles, and fig. 4 shows the results for π-mesons, protons and for deuterons and tritons taken together without discrimination. The results are plotted on a logarithmic scale. The term

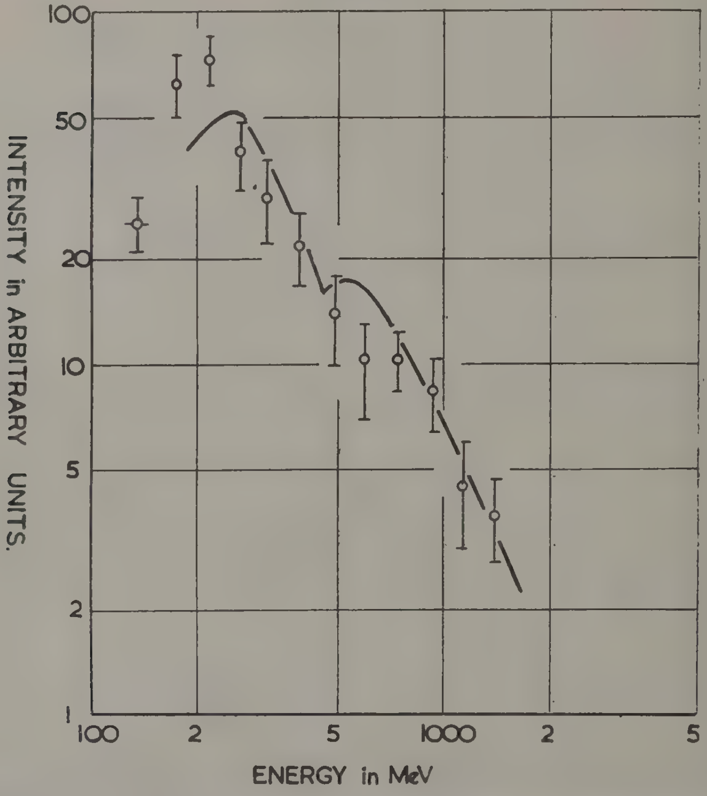
Fig. 4.



Differential distribution in kinetic energy for the various types of particles. The point at 1·6 MeV. for mesons, is for π^--particles only and is calculated from the observed number of mesons which are ejected from stars and produce nuclear disintegrations at the end of their range. To obtain the number of particles per star per MeV. interval multiply by $\sim 2{\cdot}5 . 10^{-5}$.

π-mesons, in this instance, refers to all particles less massive than protons, but the arguments advanced above give us great confidence that the great majority of the points on or near curve π (fig. 1), in the energy interval from 150 to 1500 MeV., represent π-mesons. The point in fig. 4, corresponding to a π-meson of energy 1·6 Mev., is based upon the distribution in the values of the residual range of the slow ejected π-particles which reach the end of their range in the emulsion and produce "stars". In determining this point, allowance has been made for the fact that only 72 per cent of the π^--mesons produce observable stars in photographic emulsions (Adelman and Jones 1949), and suitable geometrical corrections for "loss" have also been applied. The energy spectra have been "cut off" at energies corresponding to a mean angle of scattering of 0·022° per 100μ, since the nature of the fastest "shower" particles is ambiguous.

Fig. 5.



Differential distribution in total energy of the mesons ejected from stars. The curve drawn is the spectrum calculated by Sands.

In fig. 5, the data have been plotted in the form of a differential spectrum showing the total energy, E, of the π-mesons at their points of creation. The spectrum may be represented, to within the experimental errors, by a power law of the form $dE/E^{1\cdot5}$, in the energy region from 250 to 1400 MeV

We have found no appreciable difference in the energy spectra of the mesons produced in stars with different numbers of heavy branches. This suggests that the energy spectrum does not depend strongly on the nature of the nucleus in which the mesons are created. It follows that the observed energy spectrum shown in fig. 5 is similar to that of the mesons created in the free atmosphere at a depth of 50 gm. cm.$^{-2}$. This altitude is near the main meson production layer of the atmosphere. It is therefore of interest to compare our observations with the results of Sands (1950) who has deduced the energy spectrum of the mesons at their points of production from (*a*), measurements of the intensity of the slow μ-mesons at different depths in the atmosphere; and (*b*), the energy spectrum at sea level.

The π-mesons created in nuclear interactions taking place in the atmosphere normally decay in flight to produce μ-mesons. It may be assumed that, in the decay, the μ-mesons are emitted at random in the spatial coordinate system of the π-particles. Since the velocity of the ejected μ-meson in the centre of mass system is small compared with that of the moving π-meson, the μ-mesons appear in the laboratory coordinate system with essentially the same velocity as that of the π-particles. Thus in order to transform Sands' curve to compare with our data in fig. 5, a simple correction for the difference in mass of the π and μ-mesons need only be applied, and the rest mass energy of the π-meson then added. Sands' data is shown by the solid line in fig. 5. It can be seen that two distributions are in satisfactory agreement.

There are indications of an arrest point in our experimental distribution which corresponds roughly to the dip at 450 MeV. in Sands' curve. The statistical errors associated with our values are too large, however, to allow us to regard the observation as significant.

At kinetic energies greater than 10 mc.^{2}, where m is the rest-mass of the mesons, Sands' results indicate that the energy spectrum follows a power law of the form $dE/E^{2\cdot5}$. The results shown in fig. 4 indicate that the energy spectrum of the ejected protons also follows a similar power law, for values of the kinetic energy of the order of 1000 MeV. In order to determine the relative numbers of mesons and protons that have a scattering of less than $\cdot22^\circ/100\,\mu$, we have extrapolated the proton spectrum following a power law of the form $dE/E^{2\cdot5}$, and the meson spectrum from Sands' curve. There is a satisfactory agreement between the total number of particles expected on the basis of these extrapolations and the numbers observed.

The present results indicate that of the particles, with ionization less than four times the minimum value (44 grains per 50μ), which originate in nuclear disintegrations, $36\pm2\cdot5$ per cent are π-mesons, $51\pm2\cdot9$ per cent protons and $13\pm1\cdot5$ per cent deuterons or tritons. For "shower" particles with ionization less than 16 grains per $50\,\mu$, the figures are 79 ± 6 per cent, 17 ± 3 per cent and $4\cdot0\pm1\cdot3$ per cent respectively.

NUCLEAR INTERACTIONS OF THE SHOWER PARTICLES.

It has been pointed out in Part II. that a shower particle, created in the disintegration of one of the atoms of the emulsion, is sometimes observed to make a nuclear collision. We call such phenomena double stars (Pls. IX., X. and XI.). It is possible to measure the mean free path for these interactions, and since it is now known that 80 per cent of the shower particles are mesons, almost certainly π-particles, observations of this type give important information about the strength of their interaction with nucleons.

Method I.

The simplest approach is to make a direction comparison between the total length in the emulsion, ΣL, of the tracks of shower particles of length greater than $3000\,\mu$, and the number of nuclear interactions which they produce. We can thus confine attention to those shower particles which can be shown by scattering measurements to be mesons.

In an examination of 48 c.c. of emulsion, 241 mesons, ejected from stars with kinetic energies in the interval 150–1500 MeV., and each with a length of path in the emulsion greater than $3000\,\mu$ have been identified. The total length of the tracks is 200 cm., and six of them produce a second nuclear disintegration. A meson must, however, have travelled at least $3000\,\mu$ in the emulsion in order to be identified. If it makes a nuclear interaction before traversing this length it will be excluded from the statistics both of the nuclear interactions and of the number of tracks. The effective length we have to consider is therefore: $200-(241\times 0{\cdot}3)=128$ cm., and the average length of path per disintegration is 21 ± 9 cm. This corresponds to the passage through a thickness of emulsion of mass 82 ± 35 gm. cm.$^{-2}$, a value which may be compared with that to be expected if the atoms of the emulsion present a geometrical target area to the incident particle; viz, 90 gm. cm.$^{-2}$. (George and Jason 1949.)

Method II.

A similar calculation can be made for all "shower" particles of length greater than $400\ \mu$. This value of the minimum length allows the determination of the grain density of the interconnecting track to be made with an accuracy of at least 15 per cent. Figs. 6 and 7 show the distribution in values of the grain density of all interconnecting tracks. Tracks of a sufficient length to permit their identification are distinguished by cross marking. In some cases the identification was found to be possible even though the track was shorter than $3000\,\mu$. The corresponding distribution for all measured tracks arising from stars is shown reduced by a factor 10 in the same figures. The distributions in grain density for the tracks of particles of different types are given in fig. 8.

The mean free path for the shower particles may be calculated in the manner described in the Appendix. A value of 26 ± 7 cm. was obtained for the mean thickness of emulsion traversed before a nuclear interaction occurred. This corresponds to a thickness of emulsion of mass 102 ± 27 gm. cm.$^{-2}$ and is also indistinguishable from that corresponding to the geometrical cross section.

Fig. 6.

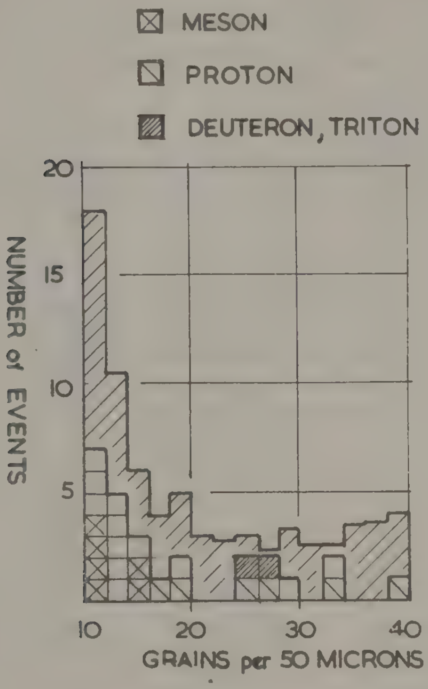
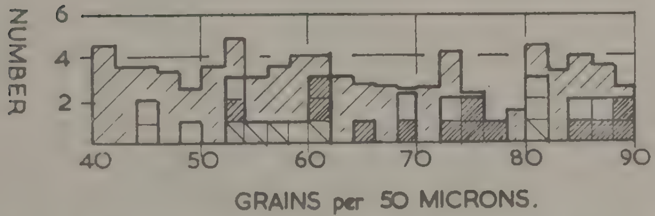


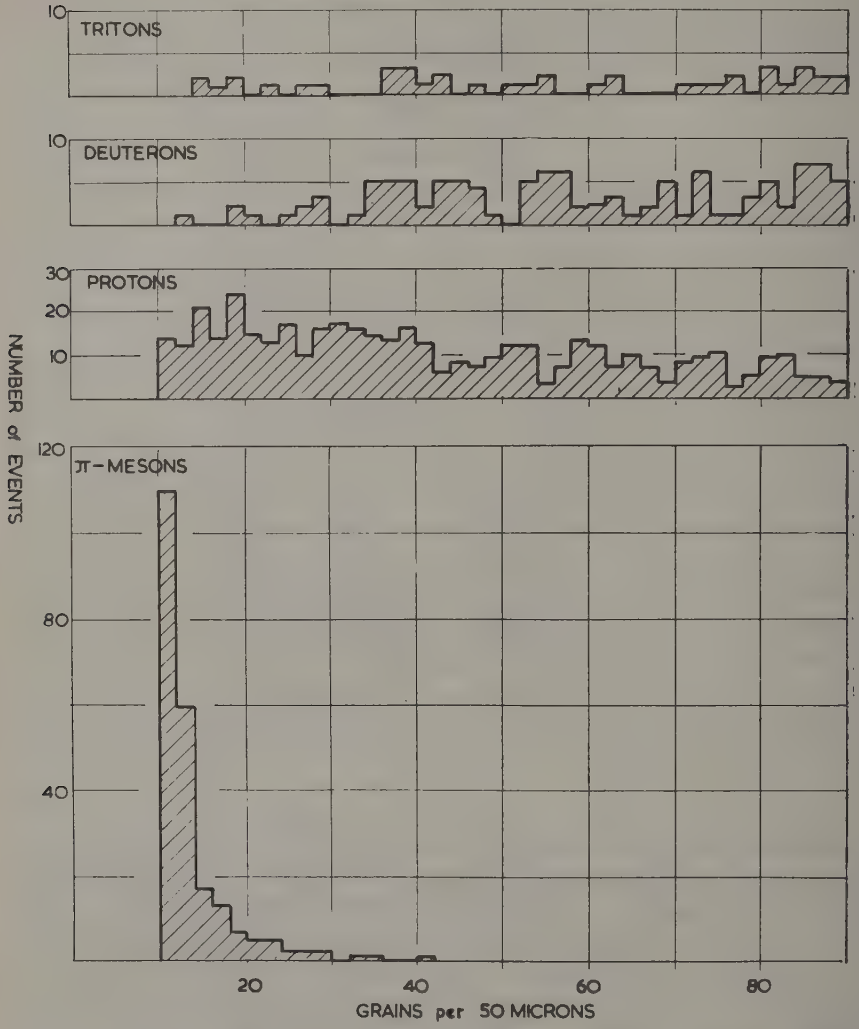
Fig. 7.



Distribution in grain-density of the tracks of particles causing secondary disintegrations. The corresponding distribution for all tracks ejected from stars is, reduced by a factor of 10, indicated by the shaded area.

We show, in a later section, that the cross section for protons with a kinetic energy between 100 and 600 MeV. is approximately geometrical. It is reasonable to assume that protons of higher energy, that is, those which are counted as shower particles, have a similar cross section. It therefore follows from the above result that the average cross section for π-mesons also corresponds to the geometrical value.

Fig. 8.



Distribution in grain-density of the tracks of various types of particles ejected from stars. To obtain the number of particles per star multiply by $\sim 1{\cdot}6 \,.\, 10^{-4}$.

Comparison with Other Work.

The value given above for the collision mean free path of the shower particles, and therefore of π-mesons, is less than that obtained by other authors. A summary of their results is given in Table I. Since the interaction of shower particles has been studied in different materials, we include in the table the ratio of the observed mean free path to the geometrical value for each determination.

All the experiments with counters and cloud chambers, however, suffer from the defect that a shower particle cannot be followed and observed continuously. The work with cloud chambers has been carried out using lead plates of various thicknesses in the chambers. The majority of the values obtained for the collision mean free path have been corrected to allow for those nuclear interactions which are contained completely within the lead plates and thus escape observation.

TABLE I.

Authors	Method	Result	$\frac{\text{Observed M.F.P.}}{\text{Geom. value}}$	Energy range (Mesons)
Piccioni	Counters at mountain altitudes	1200 gm.cm.$^{-2}$ Fe	14·0	>400 MeV.
Fretter	Cloud chamber with Pb plates at mountain altitudes	750 gm.cm.$^{-2}$ Pb Uncorrected	4·7	>150 MeV. $n_s \geq 2$
Lovati, Mura, Salvini and Tagliaferri	Cloud chamber with Pb plates at mountain altitudes	300±100 gm.cm.$^{-2}$ Pb Corrected	1·9	>150 MeV. $n_s \geq 2$
Brown and McKay	Cloud chamber with Pb plates at mountain altitudes	316±70 gm.cm.$^{-2}$ Pb Corrected	2·0	>150 MeV.
Butler, Rosser and Barker	Cloud chamber with Pb plates at sea level	400 gm.cm.$^{-2}$ Pb Uncorrected 200 gm.cm.$^{-2}$ Pb Corrected	2·5 1·4	Average energy of shower ∼7000 MeV.
Harding and Perkins	Photographic plates exposed under ice	120 gm.cm.$^{-2}$ Ice	2·0	>100 MeV.
Camerini, Fowler, Lock and Muirhead	Electron-sensitive plates exposed at high altitudes	100 gm.cm.$^{-2}$ emulsion	1·1	>150 MeV.

Table II. gives details of the type of star produced by the interaction of shower particles with the atoms of the emulsion. It can be seen that in general the stars produced are quite small, and this is confirmed by studies of the stars produced by identifiable fast π-mesons. We suggest therefore that the correction factors applied to the results of the cloud chamber work may in fact be too small. However, our results are not significantly different from the more recent results of Butler.

TABLE II.

Primary star N_H+N_S	Secondary star N_H+N_S	Length of interconnecting track	Grain density in grains per 50 μ	$\bar{\alpha}^\circ$ per 100 μ	Nature of interacting particle
$10+1_p$	$4+0_p$	3,800 μ	14·9	0·191	Meson
$17+5_p$	$6+0_p$	700 μ	13·3	—	Unidentifiable
$14+1_p$	$9+0_p$	4,600 μ	10·1	0·025	Meson
$16+1_p$	$4+0_p$	750 μ	14·5	—	Unidentifiable
$8+2_n$	$1+0_p$	3,000 μ	15·0	0·21	Meson
$3+4_p$	$5+1_p$	10,600 μ	12·2	0·0087	Unidentifiable
*$18+5_n$	$5+0_p$	10,600 μ	10·4	0·049	Meson
$6+5_n$	$16+0_p$	5,800 μ	12·4	0·0215	Unidentifiable
$9+1_n$	$5+0_p$	13,000 μ	11·7	0·032	Meson
$4+4_p$	$4+0_p$	450 μ	13·4	—	Unidentifiable
† $0+7_p$	$8+5_p$	1,200 μ	11·7	—	Unidentifiable
$6+6_p$	$14+3_p$	1,500 μ	12·1	0·017	Unidentifiable
$18+15_p$	$17+1_p$	1,700 μ	11·4	—	Unidentifiable
$9+1_p$	$5+0_p$	1,750 μ	10·0	—	Unidentifiable
$8+1_p$	$1+0_p$	11,200 μ	10·1	0·068	Meson

* See Plate X. † See Plate XI.

NUCLEAR INTERACTIONS OF PROTONS, DEUTERONS AND TRITONS.

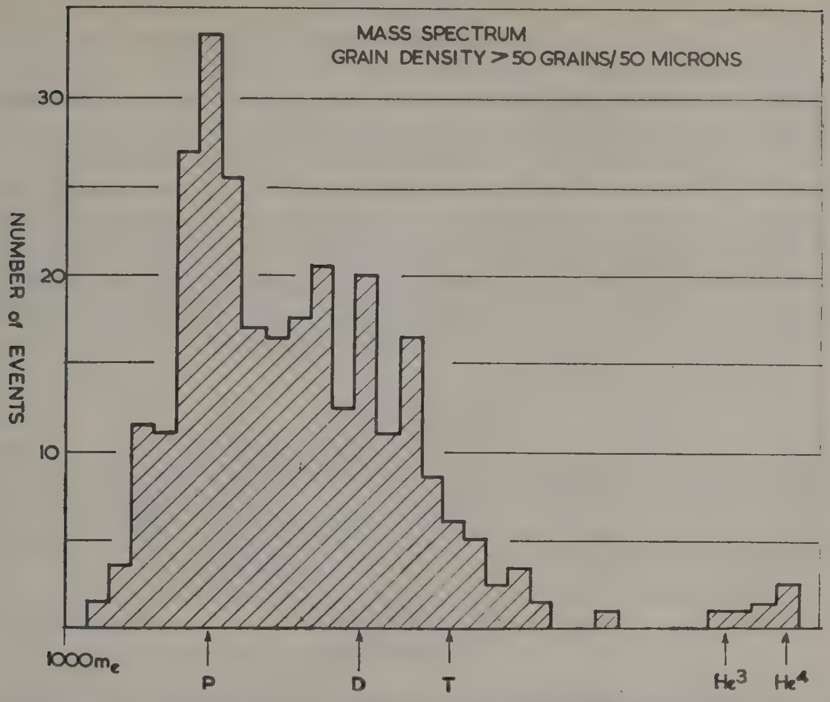
The same method of analysis which has been applied above to shower particles may be used to obtain a value for the collision mean free path of protons, deuterons and tritons, within specific energy ranges. Fig. 9 shows the mass spectrum of particles with a grain density greater than 50 grains per 50μ. It can be seen that it is possible to make a distinction between protons and deuterons, but not between deuterons and tritons.

The results obtained for the mean free path are summarized in Table III.* The figures given in column 6 of the table have been determined in the following way. We have determined the ratio of protons to deuterons plus tritons, in the different grain-density intervals, for the identified interconnecting tracks. The unidentified tracks have then been attributed to the different types of particle in accordance with this ratio. The fact that the mean free path for deuterons plus tritons is widely different from that obtained for protons, indicates that the distinction between protons and deuterons is in fact valid.

---

* The figures for shower particles are included for purposes of comparison.

Fig. 9.



Distribution in mass of the particles ejected from stars with grain-density greater than 50.

TABLE III.

Mean free path of secondary particles for nuclear interaction.

Type of particle	Energy range in MeV.	Grain density in grains per 50 μ	Effective path length considered	Number of interacting particles	Mean free path	Volume of emulsion examined
Identifiable Mesons	150–1500	10–16	128 cm.	$6 \pm 2{\cdot}5$	21 ± 9 cm.	48 c.cs.
Shower particles	150–5000 (Mesons) 600–5000 (Protons)	10–16	389 cm.	15 ± 4	26 ± 7 cm.	48 c.cs.
Protons	40–100	40–66	174 cm.	$5{\cdot}5 \pm 2{\cdot}7$	32 ± 16 cm.	87 c.cs.
Protons	100–600	10–16	291 cm.	$9{\cdot}6 \pm 3{\cdot}3$	30 ± 10 cm.	48 c.cs.
Deuterons plus tritons	75–250	40–76	127 cm.	$10{\cdot}7 \pm 3{\cdot}9$	$12 \pm 4{\cdot}5$ cm.	87 c.cs.

APPENDIX.

The determination of the mean free path in the emulsion between nuclear collisions of particles of different types can be made in the following way. Observations on the directions of motion of the shower particles show that they are distributed nearly isotropically in the lower hemisphere (Davies *et al.* 1950). We assume, as a sufficient approximation, that the distribution is truly isotropic, and that the same is true for protons, deuterons and tritons. The tracks can be regarded as rectilinear and it is then easily shown that the number $N(L)dL$ of which the true length in the emulsion lies between L and $L+dL$ is given by

$$N(L)\,dL=N_0\frac{D}{2L^2}dL \quad \text{for} \quad L>D, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where N_0 is the total number of particles ejected from stars, within the energy interval under consideration, and D is the thickness of the emulsion. The average length in the emulsion $\bar{L}$ of the tracks for which the values L lie between the limits L_1 and L_2 is given by :

$$\bar{L}=\frac{\log_e L_2/L_1}{1/L_1-1/L_2}$$

for $D=400\,\mu$, $L_1=400\,\mu$ and $L_2=20{,}000\,\mu$ this yields $\bar{L}=1596\,\mu$. We have imposed an upper limit of $20{,}000\,\mu$ for convenience of analysis.

In 48 c.c. of emulsion the number of tracks, of grain-density less than 16 grains per $50\,\mu$, and with a length in the emulsion greater than $3000\,\mu$, which have been measured, is 349. The calculated total number N_0 is then 5235 and the number greater than $400\,\mu$ in length is half this value, namely 2617·5.

From (1) the number greater than $20{,}000\,\mu$ in length is

$$\frac{N_0D}{2\times 20{,}000}=52{\cdot}35.$$

Thus the number between $400\,\mu$ and $20{,}000\,\mu$ in length is 2565 and the average effective length is $1596-400=1196\,\mu$, since the first $400\,\mu$ are not included in the analysis. To this we must add the contribution up to $20{,}000\,\mu$ for those tracks which have a length greater than this value. This is given by $52{\cdot}35\times(2{\cdot}00-0{\cdot}04)$ cm. Further, we have assumed an emulsion infinite in extent in the above calculations. Simple geometrical considerations show that, for the plates used, the total path length must therefore be reduced by 5 per cent. Thus the total effective path length $=\Sigma L=[(2565\times 0{\cdot}1196)+(52{\cdot}35\times 1{\cdot}96)]0{\cdot}95=389$ cm. The corresponding number of shower particles observed to make a nuclear disintegration is 15. Thus the collision mean free path of shower particles is 26 ± 7 cm.

ACKNOWLEDGMENTS.

We wish to thank Professor C. F. Powell, F.R.S., for his continued interest and encouragement in the course of the work ; Dr. H. Heitler for the diagrams ; Mr. A. R. Gattiker and Miss P. Dyer for the photographs ; Miss M. Stott, Miss A. Cole and Mr. M. J. Harris for assistance with the measurements ; and the team of microscope observers of this laboratory.

ERRATA.

In a previous paper (II.) an incorrect figure of 0·06 was given for the number of mesons produced per star. Our latest results indicate that this figure is 0·6±0·1. It was further stated that a large fraction of the mesons were emitted with kinetic energies of less than 60 MeV. Our latest data shows that this fraction is actually 0·1.

REFERENCES.

ADELMAN and JONES, 1949, *Phys. Rev.*, **75,** 1468 (A).
BUTLER, ROSSER and BARKER, 1950, *Proc. Phys. Soc.*, **63,** 145 ; BUTLER 1950, Harwell Conference.
BROWN and MCKAY, 1950, *Phys. Rev.*, **77,** 342.
CAMERINI, COOR, DAVIES, FOWLER, LOCK, MUIRHEAD and TOBIN, 1949 *Phil. Mag.*, **40,** 1073 (II.).
CARLSON, HOOPER and KING, 1950, *Phil. Mag.* (in course of publication).
DAVIES, FRANZINETTI and PERKINS, 1950, *Phil. Mag.* (in course of publication).
FOWLER, 1950, *Phil. Mag.*, **41,** 169 (III.).
FRETTER, 1949, *Phys. Rev.*, **76,** 511.
GEORGE and JASON, 1949, *Cosmic Radiation* (Colston Papers), 31.
HARDING and PERKINS, 1949, *Nature*, **164,** 285.
LOVATI, MURA, SALVINI and TAGLIAFERRI, 1949, *Nature*, **163,** 1003 ; 1950, *Phys. Rev.*, **77,** 284.
PICCIONI, 1950, *Phys. Rev.*, **77,** 1, 6.
SANDS, 1950, *Phys. Rev.*, **77,** 180.

EXPLANATION OF THE PLATES.

Plate IX.

A shower particle ejected from a $7+8_p$ star is observed to produce a second nuclear disintegration of the type $3+3_p$. The interconnecting track (5) has minimum grain-density and cannot be identified owing to its short length. Track (2) of the primary star is due to a meson ($\bar{\alpha}=0{\cdot}15°$ per $100\,\mu$; grain-density$=12{\cdot}9$ grains per $50\,\mu$) ; the track (*a*) in the secondary star is also due to a meson. ($\bar{\alpha}=0{\cdot}03°/100\,\mu$; grain-density$=12{\cdot}1$ grains/$50\,\mu$).

Observer: Miss B. Hulbert.

Plate X.

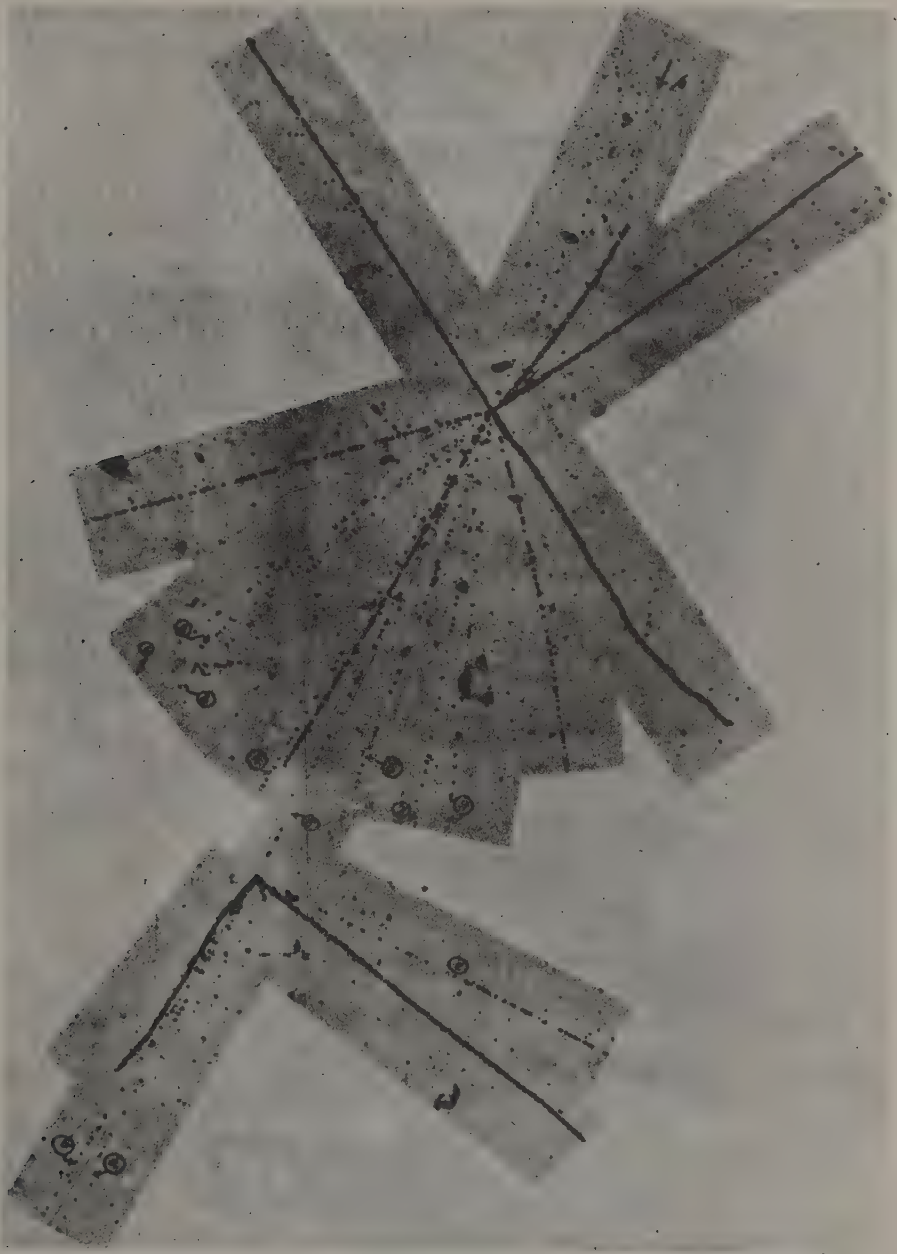
A shower particle ejected from a $18+5_n$ star is observed to produce a second nuclear disintegration of the type $5+0_p$; the interconnecting track is $10{,}600\,\mu$ long and can be identified as due to a meson ($\bar{\alpha}=0{\cdot}049°$ per $100\,\mu$; grain-density$=10{\cdot}4$ grains per $50\,\mu$).

Observer: Mrs. J. Cowie.

Plate XI.

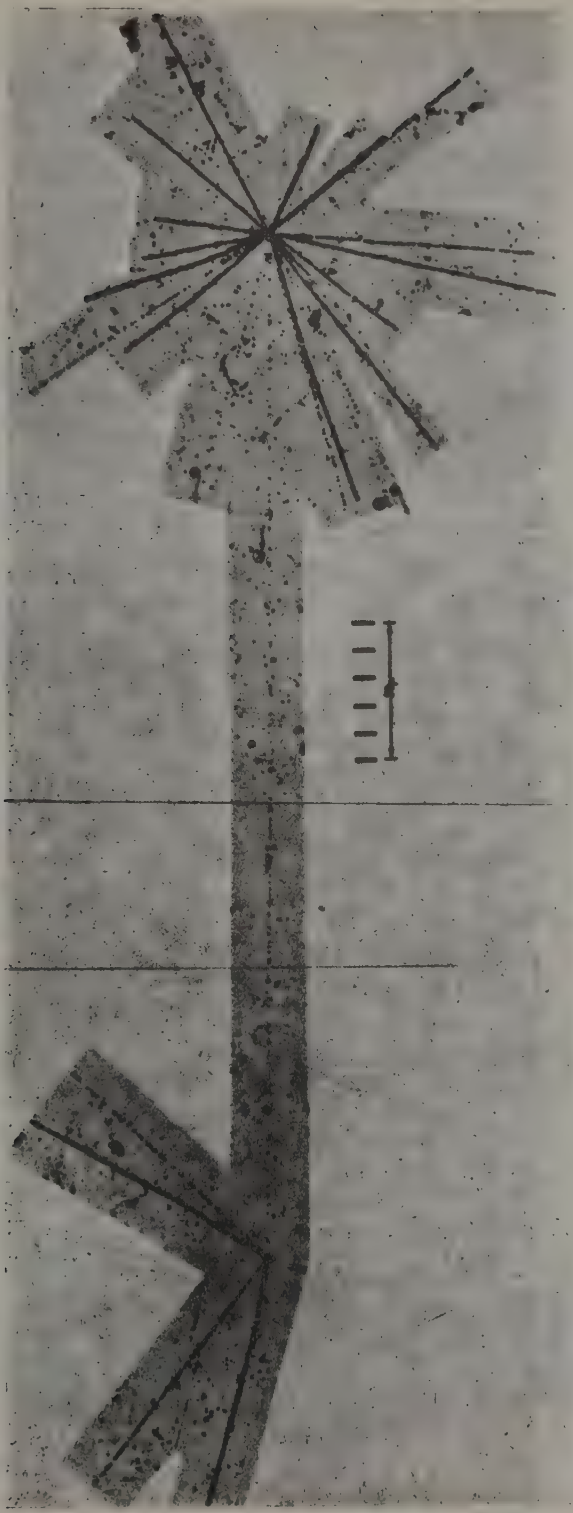
A shower particle ejected from a shower of type $0+7_p$ originating at A produces a second shower of type $8+5_p$ at B. The event was observed in the corner of a plate, and the emulsion was too distorted to allow scattering measurements to be made.

Observer: Miss R. Mitchell.

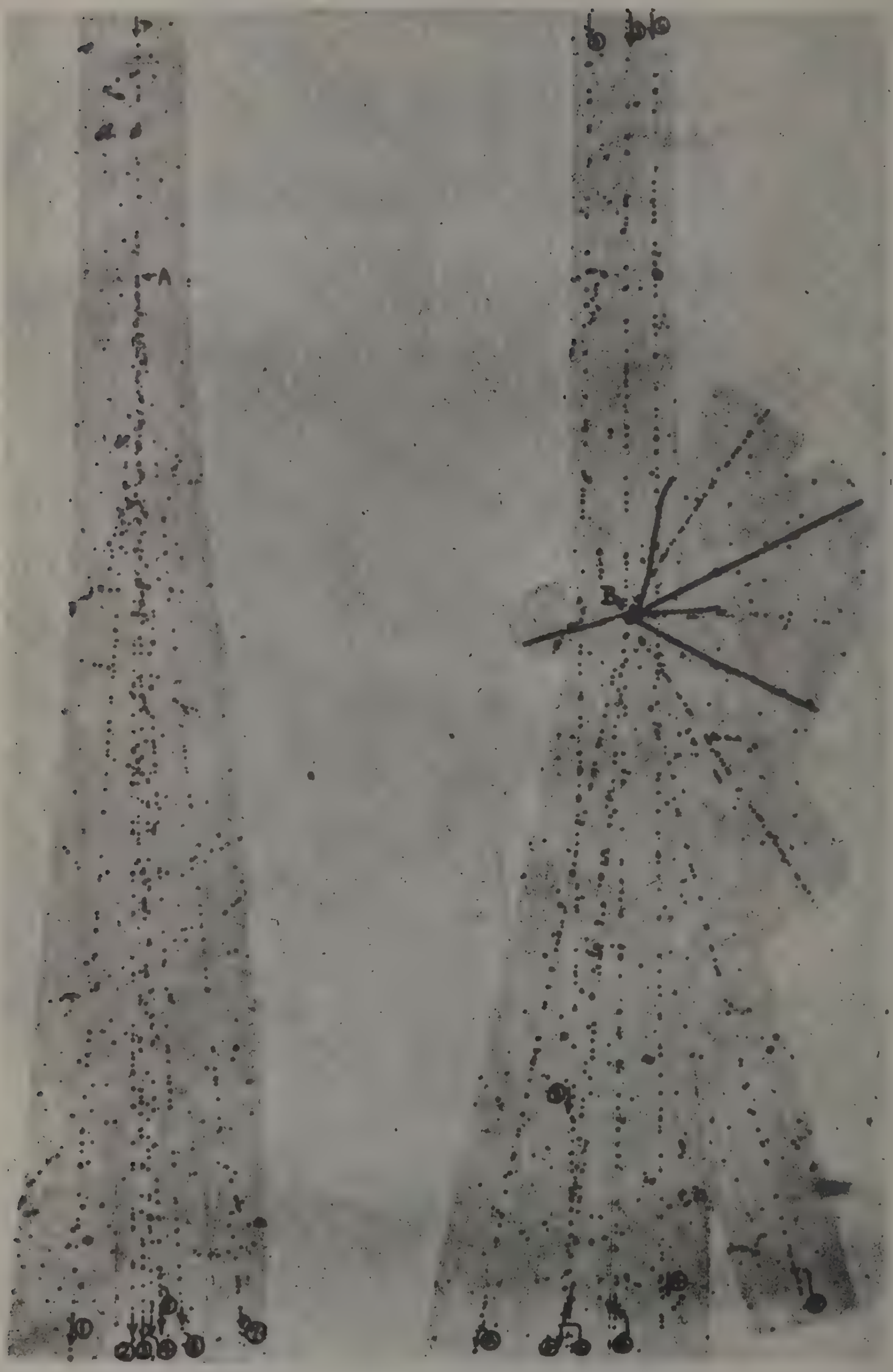


Observer: Miss B. Hulbert.

To face page 428



Observer:
Mrs. J. COWIE



Observer: Miss R. MITCHELL.

XXXVIII. *Energy Spectrum of Fragments from the Spontaneous Fission of Natural Uranium.*

By W. J. WHITEHOUSE and W. GALBRAITH *.

[Received January 31, 1950.]

ABSTRACT.

A direct comparison is made between the energy spectrum of fission fragments from the spontaneous fission and from the thermal neutron induced fission in the same foils of natural uranium. The results show that, within the limits of the experiment, the two distributions are not significantly different from one another. Calculations are made which indicate the magnitude of the distortion of the energy spectrum to be expected, and the theory is found to agree satisfactorily with experiment.

As a subsidiary result to the main experiment the rate for the spontaneous fission of natural uranium is found to be 24·2 ± 1·5 per hour per gram.

§1. INTRODUCTION.

THE possibility of the spontaneous fission of the heavy elements was forecast by Bohr and Wheeler (1939) who estimated a mean life for Uranium of the order of 10^{22} years. The phenomenon was first observed experimentally by Flerov and Petrzhak (1940, 1941). They used a multi-plate ionization chamber with active foils of U_3O_8 and obtained a value for the half-life (based on the assumption that all the spontaneous fissions were due to the U^{238} isotope) of 10^{16}–10^{17} years. This corresponds to between 2·1 and 21 fissions per hour per gram of Uranium.

Shortly after this discovery, as a result of the war, the publication of experimental results on this subject was seriously interrupted, and a review of the literature is further complicated for the present authors by the difficulty in obtaining translations of Russian papers. There have, however, been a number of studies of the neutrons emitted in spontaneous fission (Maurer and Pose 1943, Pose 1943, Chatterjee 1945, Scharff-Goldhaber and Klaiber 1946). Pose obtained a nominal half-life for uranium, with respect to neutron emission, of $3{\cdot}1\times10^{15}$ years, corresponding to the emission of about 65 neutrons per hour per gram.

N. A. Perfilov (1947) measured the spontaneous fission rates of Uranium and Thorium using the method of track counting in photographic emulsions which had been exposed to the fragments at a depth of 50 metres below the surface of the earth. An oxidizing agent was employed to remove

---

* Communicated by the Authors.

the tracks of alpha-particles from the emulsion. The value obtained for the half-life of Uranium with respect to spontaneous fission was $(1\cdot3\pm0\cdot2)\times10^{16}$ years, (16 ± 2 fissions per hour per gram). Perfilov also measured the ranges of the residual tracks, after oxidation, of the fragments from spontaneous fission, and from slow neutron-induced fission in Uranium, and in each case obtained a distribution with a single peak at about 1·0 cm. air equivalent. As he points out the fission tracks are much reduced in length by the oxidation. For this reason his results are not of any use in determining the amount of asymmetry in spontaneous fission.

A figure of 20 spontaneous fissions per hour per gram has been given for the spontaneous fission rate of U^{238} (*The Science and Engineering of Nuclear Power* 1947), but without any account of the method of measurement. It is evident, therefore, that the rate for Uranium is nearer the upper of the original limits given by Flerov and Petrzak.

It was felt that a knowledge of the distribution in energy of the fragments from spontaneous fission of uranium, as distinct from the rate at which the phenomenon occurs, would be useful in the formulation of theories of fission and it was with this object in view that the present work was undertaken.

The experiment consists essentially of a direct comparison of the energy distributions of the fragments from the spontaneous fission of uranium and from the slow neutron induced fission of U^{235}, which forms a convenient standard of comparison, under identical conditions. The latter distribution has been investigated by a number of workers (Kanner and Barschall 1940 ; Flammersfeld, Jensen and Gentner 1943 ; Jentschke 1943) but the experiments of Fowler and Rosen (1947), Deutsch and Ramsey (1945) and Brunton and Hanna (1949, 1950) * are particularly suitable for comparison with the present results, since all of them used gridded ionization chambers similar to ours.†

Brunton and Hanna give in their paper a summary of the results obtained by themselves and five previous experimenters on the distribution for U^{235}. In each case two rather broad peaks are observed. That for the light group of fragments has a most probable energy of about 92 MeV., and a width at half maximum of 12–15 MeV., while that for the heavy group of fragments has a most probable value of about 60 MeV. and a width at half-maximum of 20–25 MeV.‡ The two peaks overlap slightly, and the height of the minimum between them is approximately 1/7 to 1/10 of the higher peak (see fig. 5, curve I.).

---

* The results of the latter paper were kindly communicated to us before publication.

† In the last two experiments the energies of the two fragments were measured simultaneously, but results were also obtained when only one fragment was observed, and it is these which are used for comparison.

‡ Brunton and Hanna point out that there is a correction to be applied in their experiment for energy loss in the foil and collimator.

The difficulties associated with a similar measurement of the energies of spontaneous fission fragments arise from the very slow rate of fission. We have used a large cylindrical ionization chamber with uranium foils of total area 675 square centimetres on the inner surface of the outer electrode. Even so, in order to obtain a counting rate of about 1·4 fissions per hour, we have had to use foils of approximately 0·1 mg./cm.^{2} thickness. This distorts the observed energy distribution considerably, in a manner which will be discussed in § 3.1. Another difficulty associated with the slow counting rate is that of maintaining constant conditions in the ionization chamber itself, and in the electronic amplifiers and recording apparatus. We have overcome the difficulty associated with the chamber by using pure argon as the gas filling instead of the more usual argon–carbon dioxide mixture, and purifying it continuously, throughout the course of the experiment, by convective flow over hot calcium turnings. Our purifier worked on the same principle as that mentioned by Rossi and Staub (1949) but was simpler in design. This practice results in a longer collection time for the electrons, but does not appear to have led to any noticeable errors. The electronic apparatus was calibrated daily with a signal generator and we have good reason to believe that the error caused by random variations in the sensitivity of the chamber, amplifier and recording apparatus was not greater than ±3 per cent during the experiment. Other possible sources of error will be discussed in § 3.

In order to compare the two energy distributions under identical conditions, we simply removed part of the cadmiun shielding at different times (see § 2.2) and placed a neutron source near the chamber to induce fissions in the U^{235} present in the foils. In this way we measured approximately 1000 neutron induced fissions in three groups over a period of a week, the number of spontaneous fissions recorded during a week being about 200. This procedure was repeated during subsequent weeks until in the end some 5000 neutron induced and 1000 spontaneous fissions were obtained.

§ 2. APPARATUS

2.1. *The Ionization Chamber.*

A cylindrical form for the ionization chamber was chosen, as it eased a number of mechanical problems, particularly the construction of the grid (see fig. 1). The outer electrode is 25 cm. in diameter and 20 cm. high and is supported on porcelain insulators. The uranium foils are carried on the inner surface of this electrode and are approximately 10 cm. across, ensuring that no ionization track left the active volume of the chamber. The inner electrode consists of a cylinder 20 cm. high and 11·5 cm. diameter, and is supported on insulators consisting of perspex rings. The grid which has diameter of 15 cm. consists of 64 beryllium copper wires (24 S.W.G.) of length 25 cm. supported on half-rings of brass attached to the same insulator system as the inner electrode. It is easily demountable and quite robust. There are two guard rings to prevent transient discharges along the insulators between the grid and the inner electrode. The system of electrical connections is shown in fig. 1.

The distance between the outer electrode and the grid is about 5 cm., which is considerably longer than the track length in argon at one atmosphere of either the fission fragments or the alpha particles from uranium. The distance from the grid to the inner electrode is about 1·7 cm.

In a chamber designed to measure spontaneous fission pulses the background either from gamma rays or alpha contamination is unimportant, and no attempt has been made to minimize it.

The theory of a planar grid chamber has been given by Bunemann, Cranshaw and Harvey (1949). Their formulæ have been used to calculate the dimensions of our grid, using the following transformations between electrically equivalent cylindrical and plane chambers.

Fig. 1.

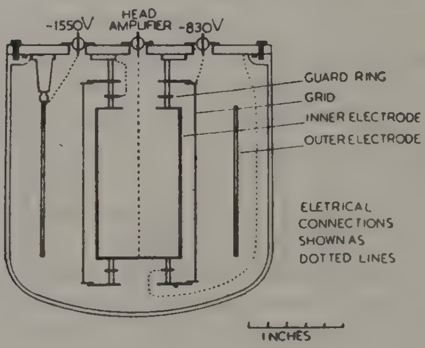


Diagram of ionization chamber.

If R_P is the radius of the inner electrode,

R_G is the radius of the grid,

R_A is the radius of the outer electrode,

then the dimensions of the electrically equivalent plane chamber are given by (see fig. 2):

$p^1 = R_G \log_e R_G/R_P$ where p^1 is the distance from the anode to the central plane of the grid.

$a^1 = R_G \log_e R_A/R_G$ where a^1 is the distance from the central plane of grid to the outer electrode.

When this system of transformation is used the physical dimensions of the grid are the same in both chambers.

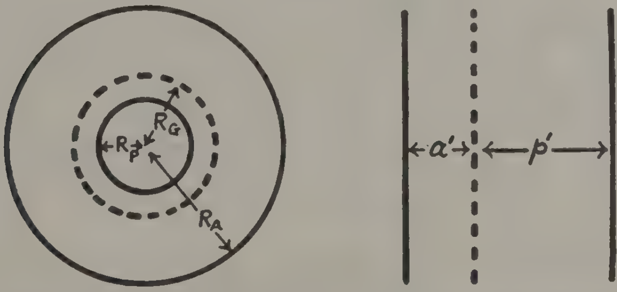
One important quantity to be calculated is the charge induced on the collecting electrode by the positive ions which, after each fission, are left in the neighbourhood of the outer electrode. The dimensions of the grid were so chosen that this was, theoretically, less than 2 per cent of the charge carried to the collecting electrode by the electrons. This rather low efficiency of the grid is tolerable because the accuracy of the measurements of the energies required in this experiment is much lower than that usually required in the measurement of alpha particle energies.

The other important condition to be satisfied is that none of the electrons should be collected on the grid wires. With our dimensions this requires that the voltages on the outer electrode (V_A), grid (V_G) and collector (V_P) should obey the relation

$$\frac{V_P - V_G}{V_G - V_A} \geqslant 0{\cdot}88. \qquad \dots\dots\dots\dots \quad (1)$$

The chamber was filled with argon (99·8 per cent pure), to a pressure of about 1·25 atmospheres, and this was continuously purified by convective flow over calcium turnings at about 400° C. The volume of the chamber is approximately 30 litres, and the rate of flow of gas was about 40 c.c. per second, so that the chamber gas was frequently renewed. After the purifier had been switched on, the conditions in the chamber, as judged from the energy and resolution of the uranium alpha-peaks, improved for about fifteen hours, and then remained constant, within our limits of accuracy, for the three months during which preliminary work and the actual counting were carried out*.

Fig. 2.



Electrically equivalent cylindrical and plane ionization chambers.

The voltages for the outer electrode and the grid were supplied by dry batteries, through simple resistance-capacity filters.

---

* We were unfortunate enough to have three failures in the power supply, early in the experiment, which resulted in some loss in gas pressure, due to the contraction of the seals of the purifier on cooling down. Apart from this there is no reason to suppose that the chamber and purifier could not have been operated continuously for a year.

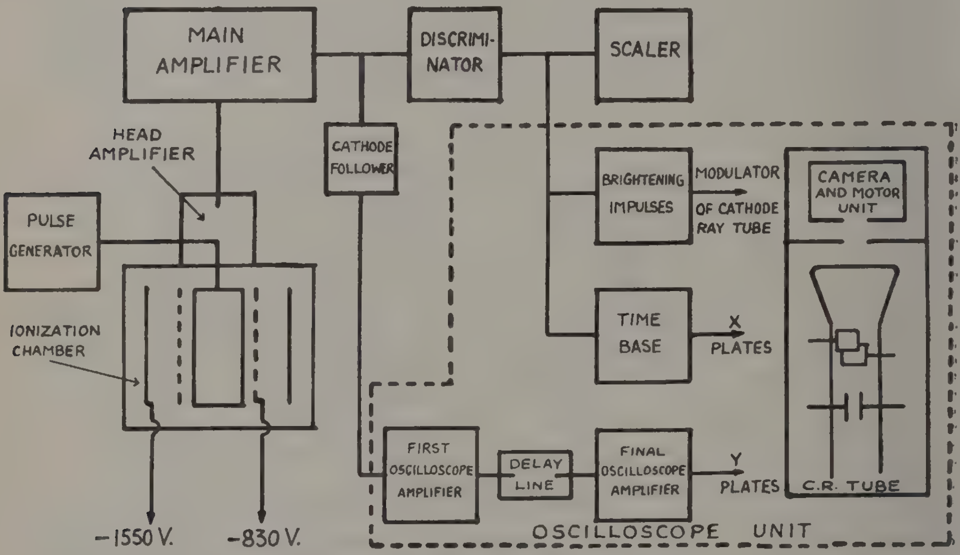
2.2. *The Neutron Shield for the Chamber.*

The chamber was surrounded by a cadmium shield (0·5 mm. thickness) around which was a paraffin shield of approximately 25 cm. thickness. A hole was cut in one of the paraffin blocks in which a neutron source could be placed. By removing a section of the cadmium shield and inserting the neutron source a satisfactory rate of neutron induced fissions could be obtained.

2.3. *The Uranium Foils.*

Eight uranium foils each 9·2 cm. square were prepared by evaporation under vacuum on to platinum. No small scale irregularities were seen when the foils were viewed under a low power microscope. The uniformity over any one foil was checked by counting the alpha-particles from sections of the foils and was constant to within ±15 per cent. The total weight of the deposit, which is presumably oxidized, and of which the chemical composition is unknown, is 71 mg., which corresponds to an average surface density of 0·1 mg./cm^2. The separate foils were found, by weighing, to have average surface density per foil between 0·09 and 0·12 mg./cm^2. The total weight of uranium present, estimated by alpha-counting, is 57±3 mg. indicating either that the films are completely oxidized or that there is some impurity present. It is fortunately not important in this experiment to have a very exact knowledge of the weight of uranium in the foils.

Fig. 3.


MAIN AMPLIFIER
DISCRIMINATOR
SCALER
HEAD AMPLIFIER
CATHODE FOLLOWER
PULSE GENERATOR
BRIGHTENING IMPULSES
MODULATOR OF CATHODE RAY TUBE
CAMERA AND MOTOR UNIT
IONIZATION CHAMBER
TIME BASE
X PLATES
FIRST OSCILLOSCOPE AMPLIFIER
DELAY LINE
FINAL OSCILLOSCOPE AMPLIFIER
Y PLATES
C.R. TUBE
−1550 V.
−830 V.
OSCILLOSCOPE UNIT

Block diagram of electronic apparatus.

2.4. *Amplifiers, Recorder, and other Electronic Apparatus.*

A block diagram of the apparatus is given in fig. 3. The head amplifier consists of one so-called " ring of three " circuit and the main amplifier of two such circuits. The circuits contained no electrolytic condensers, as these had been found to generate spurious pulses, and the grid bias voltages for the valves were derived from a resistance chain in the negative H.T. line. This modification increased the hum level of the amplifier, but not sufficiently to prevent an accurate determination of the position of the U^{234} alpha-peak. At the gain setting used for measuring the fission pulses the hum is negligible. The band width of the amplifier is controlled by " differentiating " and " integrating " condenser–resistance networks with time constants which can be varied in steps from $0{\cdot}08\,\mu$S to $8\,\mu$S. The gain is controlled by attenuators made up from high stability components.

The discriminator was a standard instrument (A.E.R.E. Type 1028). It was set at a level corresponding to an energy of 20 MeV.

The fission pulses were recorded by photographing an oscilloscope trace on 16 mm. film, and the times at which they occurred recorded by photographing a clock. All resistors in the oscilloscope amplifiers were high stability types. The fission pulses were passed through a 6μ second delay line in order to allow the time base and brightening circuits of the oscilloscope to operate. After each pulse the film was automatically moved on one frame. The interval between successive spontaneous fission pulses was sometimes several hours, in which case some fogging of the film occurred due mainly to the residual light from the filament of the cathode ray tube. This, however, was not serious.

The traces of the fission pulses were projected on to a screen for measurement.

The pulse generator, which was not a precision instrument, was coupled to the head amplifier through a small capacitor of about $1\,\mu\mu$F. It was used for intercalibrations between the discriminator setting and the cathode ray tube deflections, and also for checks on the linearity of the main linear amplifiers, and the amplifiers associated with the cathode ray tube.

It was essential to have a mains supply free from fluctuations which would have generated spurious pulses in the amplifiers. All the apparatus, except the heater for the gas purifier, was therefore supplied from a separate motor driven single phase alternator.

§3. ERRORS.

It is instructive to consider what limitations are imposed on the accuracy of the results by the conditions of the experiment.

3.1. *Energy loss of Fragments—Distortion of Energy Distribution Curve.*

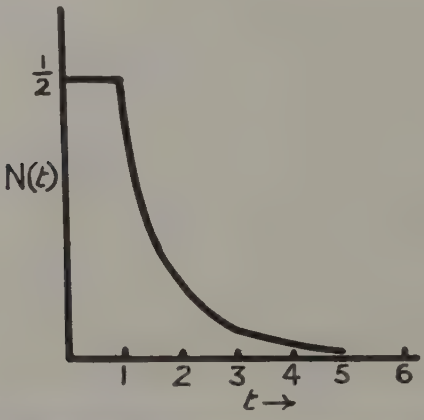
(a) *Foil thickness.*

The distribution of the paths which the fission fragments traverse inside the foil can be shown to have the form given in fig. 4 provided that the thickness of the foil is small compared with the maximum range of the fragment in the foil. If T is the path length in the foil and X_0 is the thickness of the foil, and $t=T/X_0$, the distribution is a straight horizontal line with $N(t)=\frac{1}{2}$ for values of $t \leqslant 1$, and falls off for $t > 1$ according to the formula

$$N(t)\,dt=\frac{1}{2}\frac{dt}{t^2}. \qquad (2)$$

The numerical constant $\frac{1}{2}$ is introduced to make the integral of $N(t)$ from $t=0$ to $t=\infty$ equal to unity, which corresponds with the fact that for every fission one fragment will emerge from the upper surface of the foil.

Fig. 4.



Distribution of particles originating in a foil as a function of their length of path in the foil.

The energy lost by the fragment near the beginning of its path can be calculated from the approximate formula of Bohr (1941, 1948) according to which the velocity of the fragment is

$$v=v_0(1-T/R_F), \qquad (3)$$

where v_0 is its initial velocity, T is the distance it has travelled, and R_F is an extrapolated fission range which is nearly the same as the actual range of the fragment. We have therefore

$$\epsilon=E_0\left(\frac{2T}{R_F}-\frac{T^2}{R_F{}^2}\right). \qquad (4)$$

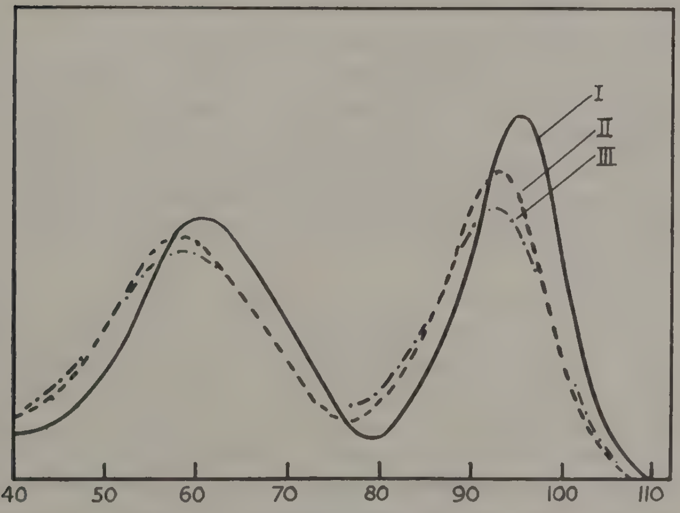
Where ϵ is the energy lost by the fragment in traversing a path T; E_0 is the initial energy of the fragment.

The value of R_F can be calculated from the range of an alpha-particle with the same initial velocity as the fragment (Bohr 1941).

We have calculated values of R_F in U_3O_8 for the light and heavy fission fragments using a relative stopping power of 4 mg./cm.^{2} U_3O_8 equivalent to 1 cm. of air (Bøggild *et al.* 1947) and have found the initial energy losses ($d\epsilon/dT$ for $T=0$) of the two fragments, in terms of the path expressed in mg./cm.^{2}. For the light fragment this is 19 MeV. per unit of path and for the heavy it is 17 MeV. This agrees reasonably well with the measurements of West (1948) who found that the fragments lose 6·5 MeV. per mm. of nitrogen, corresponding to 16 MeV./mg./cm.^{2} of U_3O_8.

In order to estimate, roughly, the distortion to be expected in our curve, we have taken the following round figures (*a*) an initial energy loss $d\epsilon/dT$ of 20 MeV. for each mg./cm^2 of film for all the fragments whether heavy or light, (*b*) a film of thickness 0·1 mg./cm^2. Thus all the fragments are assumed to lose 2 MeV. if they traverse the whole depth of the film in a direction normal to its surface.

Fig. 5.



Energy (MeV.).

Corrections to be applied to the energy distribution of fission fragments.
Curve I. Energy distribution of fission fragments from thermal neutron induced fission of U^{235} (after Brunton and Hanna).
Curve II. Curve after correction is made for energy loss in foil.
Curve III. Curve after correction both for energy loss and gain variation.

By combining equations (2), (3) and (4) and taking terms as far as the fourth power it is possible to show that the number of particles losing an energy η is given by

$$\left.\begin{array}{l} n(\eta)=\tfrac{1}{2} \text{ for } \eta\leqslant 1, \\ n(\eta)\,d\eta=\tfrac{1}{2}\,d\eta/\eta^2\{1-1/16(\epsilon_0/E_0)^2\eta^2\} \text{ for } \eta>1 \end{array}\right\} \quad . \quad . \quad . \quad (5)$$

Where $\eta=\epsilon/\epsilon_0$ and ϵ_0 (=2 MeV.) is the energy lost in traversing the film in a direction normal to its surface. Thus to a close degree of approximation the expressions for the distribution according to path lengths and to energy loss are identical in form.

When this correction is applied, curve II. of fig. 5 is obtained.

It has become a common practice in measurements of this type to collimate the particles with a perforated metal sheet, thus eliminating those which lose a large amount of energy in passing obliquely through the film. The disadvantage of this practice in an experiment such as this, where the counting rate is low, is that the collimator stops a large proportion of the particles, and necessitates the use of a much thicker source. Calculations have been carried out (Howlett and Whitehouse 1950) which show that in this experiment the use of a collimator would have increased the energy lost by the particles in the film, and a plain foil has therefore been used.

(b) *Energy lost by Particles striking the Electrode.*

In a cylindrical chamber such as ours some of the fragments lose energy at the ends of their paths by striking the curved surface of the outer electrode. This affects less than 5 per cent of the fragments, and we have estimated that the average loss in energy from this cause is less than $\frac{1}{4}$ MeV., whereas the average energy loss in the film itself is about 4 MeV. We therefore have not applied any correction for this effect.

(c) *Effect of Variations in the gain of the Amplifier and Recording Apparatus.*

Curve III. of fig. 5 shows the effect which might be expected from a random variation of the sensitivity of the apparatus. In order to simplify the computation we have assumed that the measurements for all the fragments regardless of their initial energy are subject to a random error with a standard deviation of ± 3 MeV.

This figure has been chosen as the most probable value of the spread in our results as a result of the combination of the drift of the apparatus over a period of time (§ 3.2) and of the differential sensitivity to pulses of different rise time (§ 3.3).

Owing to the rough nature of the corrections described above we do not invite detailed comparison between our results and the theoretically distorted curves of fig. 5. In fig. 6, however, we plot our experimental results for U^{235} along with curve III. of fig. 5. Our approximate energy calibration shows our results for the energies of the two peaks to be within experimental error of those given in the introduction, but in drawing fig. 6, we have adjusted the curves so that the total number of fissions is the same in each, and the positions of the minima between the peaks coincide.

It will be seen that the two main sources of error in the experiment do provide an adequate explanation for the difference between the shape of our curve and that of Brunton and Hanna.

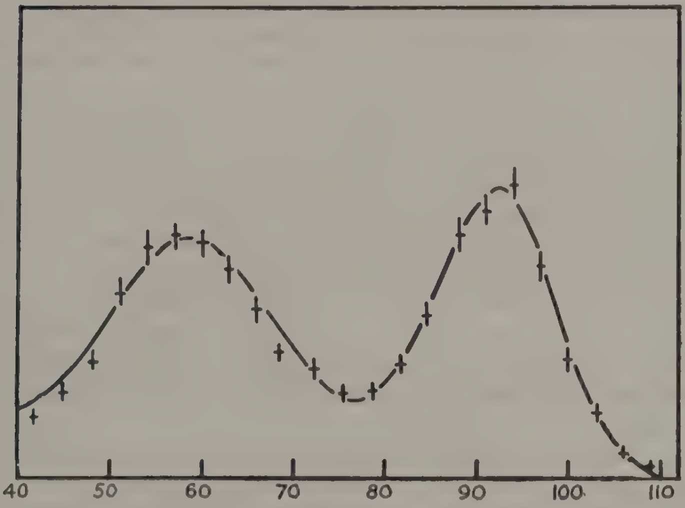
3.2. *The Drift in Gain.*

A daily check was made of the linearity of the amplifiers and their overall gain, by means of the signal generator and the zero error of the discriminator was determined at the same time. It was decided to use the discriminator voltage as a standard and to refer all the pulse heights to it.

We have two independent checks of the constancy of the apparatus :

(*a*) the height of the U^{234} alpha-peak. This was measured every second day using, of course, a different gain setting on the amplifier, and found to be nearly constant on the discriminator scale. The greatest deviation from the mean value, over three months was ±2·2 per cent.

Fig. 6.



Approximate energy scale (MeV.).

Comparison of present results for fission fragments from thermal neutron induced fission of U^{235} with corrected curve.

(*b*) It is also possible to maintain a general check on the steadiness of the apparatus from the average of the heights of the fission pulses.

If we regard the high energy peak alone as a statistical population, and apply simple sampling theory (neglecting the difference between the shape of the observed peak and the normal error curve) we can assume that the mean energy of n fissions will have had a standard deviation of approximately $\sigma/\sqrt{(n)}$, where σ is the standard deviation of the whole population. The value of σ is of the order of 10–15 MeV ; consequently the mean energy of any hundred consecutive fission pulses in the high energy group should have a standard deviation of about 1 to 1·5 MeV. If the error due to slow drift in the apparatus is much greater than this, it will show up in the mean energies of groups of fragments. Exactly similar considerations apply to the mean energies of the fissions in the low energy group.

We have calculated the mean energies of the high and low energy fragments in groups of five-hundred for neutron induced fissions and of two hundred for spontaneous fissions. We find in both cases that the standard deviation of the mean of the groups is about 2 MeV., indicating that the error due to random slow drift could not have been larger than this.

3.3. *The Electron Collection time and the Resolving time of the Amplifier.*

Owing to the comparatively slow rate of electron drift in pure argon, the large dimensions of the chamber, and the moderate collecting voltages applied, it was considered necessary to check that the resolving time of the amplifier (8 μS) was sufficient to allow complete collection of the electron pulses. This was done by experiments on the U^{234} alpha-peaks and on the energy distribution of fission fragments from neutron induced fission of U^{235}.

(*a*) It was found that it was necessary to reduce the time-constant of the amplifier from eight to five microseconds before there was any observable diminution of the α-pulse height.

(*b*) The outer electrode potential could be reduced from -1550 volts to -830 volts, and the grid voltage from -830 volts to -450 volts, without any diminution in the height of the α-pulse.

(*c*) With the reduced electrode voltages given above, the energy distribution of some 240 neutron induced fission pulses from U^{235} showed no significant difference from that obtained with the chosen operation voltages. We concluded then that the resolving time of the amplifier was adequate.

Another small effect of the electron collection time, quite unconnected with the resolving time of the main amplifier, has been noticed.

By direct calibration with voltage pulses fed into the head amplifier we have found that the amplifier pulse height measured at the discriminator is proportional to the original voltage, and is not noticeably affected by the time of rise of the pulse within quite wide limits.

The heights of the traces on the oscilloscope do, however, decrease by about 3 per cent when the rise time is reduced from 3 μs. to a fraction of a microsecond, an effect which we attribute to the delay line. Since the collection time of electrons in our chamber depends upon the angle of the track, a slight spread in the measured energy values of fission fragments will be introduced, in addition to that caused by random drift in the gain of the apparatus.

3.4. *The background Rate of Real or Spurious Pulses.*

The neutron induced fission rate, with the Ra-α-Be source in position, was about one hundred per hour, while the spontaneous fission rate was about 1·4 per hour. It follows that the background of spontaneous fissions during a period of measurement on neutron-induced fissions was too small to affect the observed distribution of pulse sizes.

By placing the neutron source outside the paraffin and comparing the count rates with and without the cadmium we were able to show that the percentage of spontaneous fissions which could be ascribed to neutrons penetrating the shield is probably negligible. Even if an unfavourable interpretation is placed upon the statistics of the results, the percentage would not be greater than of the order of 2 per cent.

Wheeler (1949) has pointed out that the capture of a negative μ meson by a uranium atom will in almost every case lead to the fission of the uranium nucleus, but that the number of fissions so induced will be negligible in comparison with the spontaneous rate.

The background of spurious counts was very low in comparison to the observed spontaneous fission rate. When the foils were removed from the chamber only two counts were recorded in 160 hours, and both these were too small to have been confused with fission pulses.

3.5. *Coincidences of Alpha-particles and Fission Pulses.*

Examination of the photographs of the trace of the cathode ray tube reveals the alpha-particles as small pulses along the time base. It is possible to make a fairly close estimate of how close the alpha-particle and the main fission pulse would have to be before it became impossible to distinguish them from each other. In this way we estimate that less than 1 per cent of the fission pulses are in effective coincidence with alpha-particles.

The chance of sufficient alpha-particles coinciding with each other to give an alpha build-up pulse which might be mistaken for a fission is negligible.

§ 4. RESULTS.

The results of the experiment are presented in fig. 7. The full line is the energy distribution curve for neutron-induced fissions, which, as has been explained, is consistent with the results of other observers, providing that allowance is made for absorption in the foil, and for some random variation in the recording apparatus. About 4700 neutron-induced fissions were used in the preparation of this curve, and the standard errors of the points are shown.

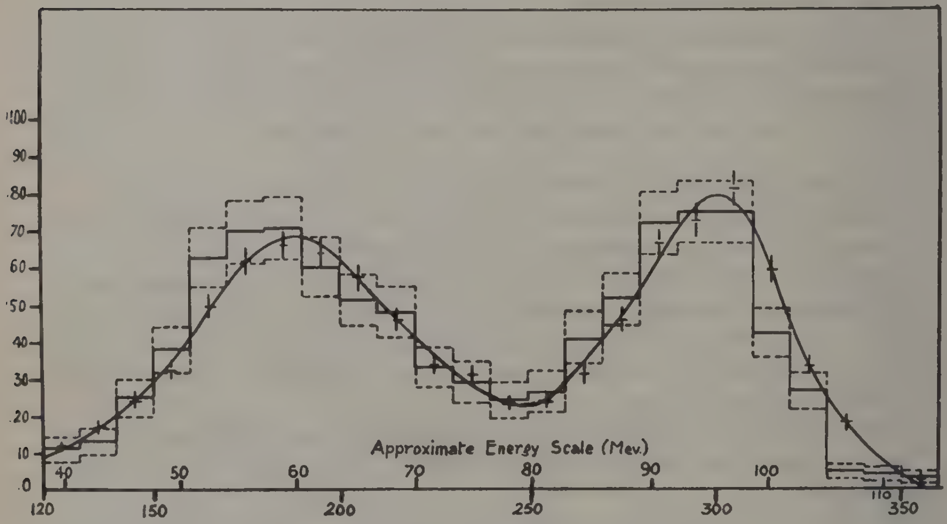
The energy distribution of 960 spontaneous fissions is shown by the histogram. The errors indicated are the square roots of the numbers of events in each cell of the histogram. The curve and the histogram have been adjusted to have the same area.

The energy scale on the diagram, derived from the height of the U^{234} alpha peak, is approximate and, in view of the errors involved in its derivation, we would not expect it to be accurate to better than 5 per cent. However, it suffices to show that our observed values for the most probable energies of the fission fragments are not inconsistent with those of other workers. An accurate absolute energy determination was not attempted, since the present experiment consists of a direct comparison between the distributions for the two types of fission.

There is no significant difference between the observed distributions for neutron-induced fissions in U^{235} and for spontaneous fissions, according to the standard χ^2 test for goodness of fit (0·05 criterion)*.

It is too much to hope, with the small number of spontaneous fissions observed, that detailed differences in the shape of the two distribution curves would be revealed, but it might be thought possible to observe

Fig. 7.



Arbitrary scale proportional to energy.

Comparison of thermal neutron induced fission pulses of U^{235}, and spontaneous fission pulses of natural uranium.

Full line curve—U^{235} fission with thermal neutrons.

Histogram—spontaneous fission.

some difference in the positions of the peaks representing the most probable energies for the fragments. We have carried out a simple test of this point by moving a tracing of the continuous curve along the energy axis, over the histogram, and carrying out χ^2 tests for different positions. We find that we get the best fit if we assume that the whole distribution for the spontaneous fissions is about 1·5 MeV. below that for the neutron-induced fissions. The χ^2 test shows, however, that the two curves are consistent for any downward displacement of the spontaneous fission distribution between 0 and 3 MeV.

---

* This means that the probability of the difference between the two distributions being due to chance is greater than 0·05.

The present experiment is not particularly well adapted to an accurate measurement of the spontaneous fission rate but a combination of the observed frequency of the spontaneous fissions ($1\cdot38 \pm 0\cdot04$ per hour) and the mass of uranium in the foils (57 ± 3 mg.) gives a figure of $24\cdot2 \pm 1\cdot5$ per gram of natural uranium per hour.

ACKNOWLEDGMENTS.

This work is published by permission of the Director of the Atomic Energy Research Establishment. The authors wish to express their thanks to Messrs. Wells and Godfrey who designed and constructed the oscilloscope unit, to Messrs. Kandiah and Gillespie for the design and modification of the head and main amplifier circuits, to Mr. Robins and his group for the preparation of the uranium foils, and to Dr. E. Bretscher, Head of the Nuclear Physics Division at A.E.R.E.

REFERENCES.

BØGGILD *et al.*, 1947, *Phys. Rev.*, **71,** 281.
BOHR, N., 1941, *Phys. Rev.*, **59,** 270 ; 1948, *Det. Kgl. Danske Vidensk Selsk.*, *Mat. Fys. Medd*, 18, No. 8.
BOHR and WHEELER, 1939, *Phys. Rev.*, **56,** 426.
BRUNTON and HANNA, 1949, *Phys. Rev.*, **75,** 990 ; *Can. Journ. Res.* A. In press.
BUNEMANN, CRANSHAW and HARVEY, 1949, *Can. Journ. Res.*, A **27,** 191.
CHATTERJEE, 1945, *Indian Journal of Physics*, **19,** 211.
DEUTSCH and RAMSEY, 1945, LADC. 257.
FLAMMERSFELD, JENSEN and GENTNER, 1943, *Zeit f. Physik*, **120,** 450.
FLEROV and PETRZAK, 1940, *J. Phys. U.S.S.R.*, No. 4–5, **3,** 275 ; 1941, *Ibid.*, No. 3, **4,** 283.
FOWLER and ROSEN, 1947, *Phys. Rev.*, **72,** 926.
HOWLETT and WHITEHOUSE, 1950, A.E.R.E. Report N/R. 473.
JENTSCHKE, 1943, *Zeit. f. Physik*, **120,** 165.
KANNER and BARSCHALL, 1940, *Phys. Rev.*, **57,** 372.
MAURER and POSE, 1943, *Zeit. f. Physik*, **121,** 285.
PERFILOV, 1947, *J. Expt. Theor. Phys. U.S.S.R.*, **17,** 476 ; *J. Phys. U.S.S.R.*, No. 3, **11**.
POSE, 1943, *Zeit. f. Physik*, **121,** 293.
ROSSI and STAUB, 1949, *Ionization Chambers and Counters*, McGraw-Hill, p. 109.
SCHARFF-GOLDHABER and KLAIBER, 1946, *Phys. Rev.*, **70,** 229.
Science and Engineering of Nuclear Power, 1947, Addison Wesley Press, pp. 80 and 358.
WHEELER, 1949, *Rev. Mod. Phys.*, **21,** 133.
WEST, D., 1948, *Can. J. Res.* A, **26,** 115.

XXXIX. *Scattering of 90 MeV. Neutrons by Alpha-Particles.*

By J. HEIDMANN,
Ecole Polytechnique, Paris*.

[Received January 2, 1950].

§1. INTRODUCTION.

THE purpose of this note is to report calculations of the scattering of 90 MeV. neutrons by alpha-particles, which could be useful for the investigation of collisions between nucleons and nuclei built up of α-particles. Using the Born approximation and the nucleon–nucleon interaction introduced by R. Serber, we have calculated :

(i.) The inelastic scattering, especially for the collision in which a deuteron is produced by stripping.
(ii.) The elastic scattering, estimating the errors due to the Born approximation.

No measurements being available on scattering of 90 MeV. neutrons by helium, we attempt to deduce values for comparison with our calculations from measurements on neutron-carbon and neutron-oxygen collisions ; this is done using the alpha-particle model for these nuclei.

§2. CALCULATIONS AND RESULTS.

As it is likely that all the nuclei produced by the collisions are formed in their ground state, the following cases arise :

(i.) Elastic scattering.
(ii.) Stripping scattering in which one neutron or one proton is stripped off.
(iii.) Formation of a deuteron plus a triton
(iv.) Scattering in which the alpha-particle is broken into two deuterons.
(v.) Scattering in which in addition one of these deuterons is disintegrated.
(vi.) Complete disintegration.

As cases (iv.)–(vi.) require more energy to be given up by the incident particle, they will be less probable ; we shall thus deal only with the three first cases. Let us begin with case (iii.)

---

* Communicated by the Author.

(A) *Production of Deuterons in Neutron-α Particle Collisions.*

(a). *Calculation.*

Let 1, 2 refer to the protons, 3 and 4 to the neutrons of the alpha-particle, 5 to the incoming neutron, and suppose that in the final state the deuteron (1, 5) and the triton (2, 3, 4) are formed. The Hamiltonian can be written, in the centre of mass system,

$$-\frac{\hbar^2}{2M}\frac{5}{6}\Delta_R+H_T+H_D+H'_{TD}=E, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where **R** is the distance between the centre of mass of the deuteron and the triton, H_D and H_T the Hamiltonians of the deuteron and of the triton, and H'_{TD} the interaction terms which we shall treat as a perturbation.

We can expand the antisymmetric wave function of the system ψ_a. in terms of the eigenfunctions of H_D and H_T,

$$\psi_a=\sum_{\epsilon\epsilon'}\psi_{T\epsilon}\psi_{D\epsilon'}F_{\epsilon\epsilon'}(\mathbf{R}).$$

Then from (1) we get

$$\sum_{\epsilon\epsilon'}\psi_{T\epsilon}\psi_{D\epsilon'}\left(\frac{-\hbar^2}{2M}\frac{5}{6}\Delta_R+\epsilon+\epsilon'-E\right)F_{\epsilon\epsilon'}=-H'_{TD}\psi_a. \quad . \quad . \quad . \quad (2)$$

Taking the scalar product of both sides of (2) with $\psi_{T\epsilon}\psi_{D\epsilon'}$ we get for $F_{\epsilon\epsilon'}$ the equation

$$\left[\Delta_R+\frac{12M}{5\hbar^2}(E-\epsilon-\epsilon')\right]F_{\epsilon\epsilon'}=\frac{12M}{5\hbar^2}\sum_{\text{spin}}\int\psi^*_{T\epsilon}\psi^*_{D\epsilon'}H'_{TD}\psi_a\,d\mathbf{q},$$

where **q** represents all the internal coordinates. Using the Born approximation we replace ψ_a by the initial wave function

$$(1-P_{35}-P_{45})\chi_l\phi_\alpha\exp(i\mathbf{k}\,.\,\mathbf{r}),$$

which is correctly antisymmetrized with the forms given later for ϕ_α and χ_l. Here ϕ_α is the spatial part of the wave function of the ground state of the alpha-particle, χ_l the spin wave function, **r** the distance between the incoming neutron and the centre of mass of the alpha-particle and P_{ij} the operator exchanging both spin and space coordinates of nucleons i and j. The amplitude of the asymptotic form of $F_{\epsilon\epsilon'}$ is then

$$f_{\epsilon\epsilon'}(\theta,\phi)=-\frac{1}{4\pi}\frac{12M}{5\hbar^2}\sum\iint\exp(-i\mathbf{K}\,.\,\mathbf{R})\psi^*_{T\epsilon}\psi^*_{D\epsilon'}H'_{TD}(1-P_{35}-P_{45})$$
$$\exp(i\mathbf{k}\,.\,\mathbf{r})\chi_l\phi_\alpha\,d\mathbf{q}\,d\mathbf{r}, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where **K** is a vector, with polar angles $\theta\phi$ relative to **k** and of modulus

$$\sqrt{\{12M(E-\epsilon-\epsilon')/5\hbar^2\}}.$$

As spin functions of the triton we shall take

$$S_1=(\alpha_3\beta_4-\beta_3\alpha_4)\alpha_2/\sqrt{2}, \qquad S_2=(\alpha_3\beta_4-\beta_3\alpha_4)\beta_2/\sqrt{2}.$$

For the deuteron

$$\mathrm{R}_1=\alpha_1\alpha_5, \qquad \mathrm{R}_2=\beta_1\beta_5, \qquad \mathrm{R}_3=(\alpha_1\beta_5+\beta_1\alpha_5)/\sqrt{2},$$

and for the initial state

$$\chi_1=(\alpha_1\beta_2-\beta_1\alpha_2)(\alpha_3\beta_4-\beta_3\alpha_4)\alpha_5/2, \qquad \chi_2=(\alpha_1\beta_2-\beta_1\alpha_2)(\alpha_3\beta_4-\beta_3\alpha_4)\beta_5/2,$$

where α_i and β_i are the spin wave functions belonging to values $\frac{1}{2}$ and $-\frac{1}{2}$ of the z-component of the spin of nucleon i.

As spatial wave functions for the deuteron, triton and alpha-particle we shall take gaussian functions of the type

$$\phi=\mathrm{A}\exp\big(-\nu\sum_{i<j}\mathbf{r}_{ij}^2\big). \qquad \qquad (4)$$

With such choices ψ_a is correctly antisymmetrized.

As nucleon–nucleon potential we have

$$\mathrm{V}_{ij}=-\tfrac{1}{2}(1+\mathrm{P}_{ij}^{\chi})(1-g+g\mathrm{P}_{ij}^{\sigma})\mathrm{V}\exp(-\beta\mathbf{r}_{ij}^2),$$

which is charge-independant.

As there are no tensor forces in this potential, we can separate in (3) the integration over spin and over spatial coordinates. P_{ij}^{χ} is the operator exchanging space coordinates of nucleons i and j, while P_{ij}^{σ} is exchanging spin coordinates.

In order to make a part of the calculations applicable to the other scattering cases dealt with later, we shall write $\mathrm{H}'_{\mathrm{TD}}$ as

$$\mathrm{H}'_{\mathrm{TD}}=\sum_{n=1}^{4}\mathrm{V}_{5n}+\sum_{n=2}^{5}\mathrm{V}_{1n}-2\mathrm{V}_{51}.$$

Then decomposing $(1-\mathrm{P}_{35}-\mathrm{P}_{45})\exp(i\mathbf{k}\,.\,\mathbf{r})\chi_l\phi_\alpha$ in symmetric and antisymmetric functions of spin and space coordinates, we get for $f_{\mathrm{EE}'}$ the expression, with l, im referring to spin states

$$f_{l,\,im}(\theta,\phi)=-\frac{1}{4\pi}\frac{12\mathrm{M}}{5\hbar^2}A_{\mathrm{T}}\mathrm{A}_{\mathrm{D}}\mathrm{A}_{\alpha}\mathrm{V}(l,\,im)$$

with

$$(l,im)=-(\mathrm{S}_i\mathrm{R}_m,\,\chi_l)[3(1-g)\mathrm{J}_1+\tfrac{1}{2}(1-3g)(\mathrm{I}_2+\mathrm{I}_4)-2(2-3g)\mathrm{I}_5],$$

$$\mathrm{J}_1=(\phi_{\mathrm{F}},\,\mathrm{U}_{12}\phi_{\mathrm{I}}), \qquad \mathrm{I}_2=(\phi_{\mathrm{F}},\,\mathrm{U}_{52}\phi_{\mathrm{I}}), \qquad \mathrm{I}_4=(\phi_{\mathrm{F}},\,\mathrm{U}_{52}\mathrm{P}_{52}^{\chi}\phi_{\mathrm{I}}),$$

$$\mathrm{I}_5=(\phi_{\mathrm{F}},\,\mathrm{U}_{52}\mathrm{P}_{53}^{\chi}\phi_{\mathrm{I}}),$$

where we put for convenience

$$\mathrm{U}_{ij}=\exp(-\beta\mathbf{r}_{ij}^2),$$

$$\phi_{\mathrm{I}}=\exp(i\mathbf{k}\,.\,\mathbf{r})\phi_\alpha/\mathrm{A}_\alpha,$$

$$\phi_{\mathrm{F}}=\exp(i\mathbf{K}\,.\,\mathbf{R})\phi_{\mathrm{T}}\phi_{\mathrm{D}}/\mathrm{A}_{\mathrm{T}}\mathrm{A}_{\mathrm{D}}.$$

The differential cross section is then

$$\frac{d\sigma}{d\Omega}=\frac{2\mathrm{K}}{3k}\sum_{i,\,m}|f_{1,im}|^2,$$

that is to say,

$$\frac{d\sigma}{d\Omega}=\Gamma\tfrac{3}{4}\,|\,3(1-g)\mathrm{J}_1+\tfrac{1}{2}(1-3g)(\mathrm{I}_2+\mathrm{I}_4)-2(2-3g)\mathrm{I}_5\,|^2,$$

with

$$\Gamma=\frac{2\mathrm{K}}{3k}\cdot\frac{144}{16\times25}\cdot\frac{\mathrm{M}^2}{\pi^2\hbar^4}\cdot\mathrm{A}_{\mathrm{T}}^2\mathrm{A}_{\mathrm{D}}^2\mathrm{A}_{\alpha}^2\mathrm{V}^2.$$

The parameters of the potential were adjusted to the low energy proton–proton scattering experiments and to the binding energy of the deuteron (Hoisington *et al.* 1939) :

$$g=0\cdot211,\qquad \mathrm{V}=45\ \mathrm{MeV.},\qquad \beta=0\cdot266\ 10^{26}\ \mathrm{cm.}^{-2}.$$

The parameters of the wave functions of the deuteron, triton and alpha particle have been determined so as to give the best binding energy for these nuclei (Gerjuoy and Schwinger 1942) :

$$\nu_{\mathrm{D}}=0\cdot0716,\qquad \nu_{\mathrm{T}}=0\cdot0517,\qquad \nu_{\alpha}=0\cdot0517\ 10^{26}\mathrm{cm.}^{-2}$$

(we made the permissible approximation $\nu_{\mathrm{T}}=\nu_{\alpha}$).
Hence

$$\mathrm{A}_{\mathrm{D}}=3\cdot12\ 10^{18},\qquad \mathrm{A}_{\mathrm{T}}=1\cdot47\ 10^{37},\qquad \mathrm{A}_{\alpha}=1\cdot17\ 10^{56}\ \mathrm{c.g.s.}$$

For 90 MeV. neutrons we get

$$k=1\cdot68,\qquad \mathrm{K}=1\cdot78\ 10^{13}\mathrm{cm.}^{-1}.$$

(b). *Results.*

There is a small backwards scattering of the order of 0·1 per cent of the forwards one ; the differential cross section can then be fairly well approximated by a gaussion function

$$\frac{d\sigma}{d\Omega}=0\cdot76\exp\left(-15\cdot1\ \theta^2\right)\times10^{-25}\ \mathrm{cm.}^2.$$

in the centre of mass system (see fig. 1, curve (*b*)).

The total cross section is then

$$\sigma_{\mathrm{D}}=0\cdot16\ 10^{-25}\mathrm{cm.}^2.$$

Transforming to the laboratory system we see that the angular distribution of the deuterons has a half-half-width Γ_0 equal to 8·4 degrees. Their energy spectrum consists of a sharp line at ~70 MeV. with an half-width of about 0·6 MeV.

(B). *Elastic Scattering.*

(a). *Results.*

Using essentially the same formalism we get

$$\frac{d\sigma}{d\Omega}=\frac{9M^2}{25\pi^2\hbar^4}A_\alpha^4V^2(1-g)^2\,|\,I_1+I_2-2I_3\,|^2$$

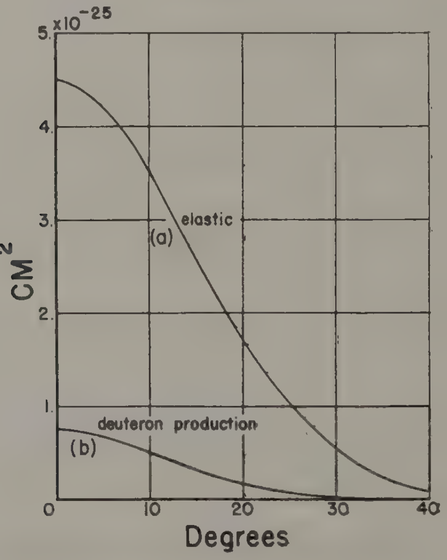
with

$$I_1=(\phi_F,\,U_{51}\phi_I),\qquad I_2=(\phi_F,\,U_{51}P_{51}^{\chi}\phi_I),\qquad I_3=(\phi_F,\,U_{52}P_{51}^{\chi}\phi_I), \quad . \quad . \quad . \quad (6)$$

where U_{ij} and ϕ_I have the same meaning as in (5) and $\phi_F=\exp(i\mathbf{K}\,.\,\mathbf{r})\phi'_\alpha A_\alpha$. The backwards scattering is small, being of the order of 0·3 per cent of the forward one; then the differential cross section can be written, in the centre of mass system,

$$\frac{d\sigma}{d\Omega}=4{\cdot}50\exp(-7{\cdot}86\,\theta^2)\times10^{-25}\text{ cm.}^2$$

Fig. 1.



(see fig. 1, curve (a)). The total cross section is

$$\sigma_{el}=1{\cdot}80\times10^{-25}\text{cm.}^2.$$

(b). *Correction to the Born Approximation.*

Consider the expansion in Legendre polynomials, of the scattered amplitude obtained in the Born approximation

$$\sqrt{\frac{d\sigma}{d\Omega}}=k^{-1}\sum_{l=0}^{\infty}(2l+1)\eta_l P_l(\cos\theta). \quad . \quad . \quad . \quad . \quad . \quad (7)$$

We may hope to get a better approximation than Born's by replacing in (7) the approximate expression η_l of the contribution of the phases η_l by the exact expression (Mott and Massey 1933):

$$(\exp(2i\eta_l)-1)/2i.$$

When using Gaussian functions, as we did for the wave functions and the radial part of the potential, the integrals defined by (6) take the simple form

$$\mathrm{I}_n=a_n \exp(b_n \cos\theta).$$

It is then possible to get for the phases η_l the formula

$$\eta_l=k\sqrt{\frac{\pi}{2}}\sum_{n=1}^{3}\frac{a_n}{\sqrt{b_n}}\,i^{-(l+\frac{1}{2})}\mathrm{J}_{l+\frac{1}{2}}(ib_n),$$

where the functions $\mathrm{J}_{l+\frac{1}{2}}$ are Bessel functions. The numerical calculation gives

$$\eta_0=0{\cdot}761, \qquad \eta_1=0{\cdot}733, \qquad \eta_2=0{\cdot}452.$$

For $l>2$ the phases are small enough for it to be possible to write

$$(\exp(2i\eta_l)-1)/2i\sim\eta_l.$$

Then the corrected cross section is

$$\tilde{\sigma}_{el}=\sigma_{el}-\frac{4\pi}{k^2}\sum_{l=0}^{2}(2l+1)(\eta_l^2-\sin^2\eta_l)=1{\cdot}60\times10^{-25}\ \mathrm{cm.}^2.$$

(C). *Stripping Scattering.*

(a). *Stripping Scattering of a Proton.*

We shall first consider the case in which the stripped nucleon is a proton.

(1). *Calculation.*

Let 1 be that proton, and denote by $\hbar\mathbf{k}'$ the momentum of the stripped proton in the alpha-particle centre of mass system, $\hbar\mathbf{k}$ and $\hbar\mathbf{K}$ the initial and final momenta of the neutron 5 in the five nucleons' centre of mass system.

The final wave function can be written

$$\psi_{\mathrm{F}}=(2\pi)^{-3/2}\mathrm{T}_i\phi_{\mathrm{T}}\exp(i\mathbf{k}'\,.\,\mathbf{R}), \qquad \ldots\ldots \quad (8)$$

where $\mathbf{R}$ is the distance between the proton 1 and the centre of mass of the triton (2, 3, 4) and ϕ_{T} is defined as in (4), and T_i is the spin wave function.

Using the same formalism as in paragraph 2, ((A) (a)), we get for the scattered intensity the expression

$$\mathrm{I}=\frac{9\mathrm{M}^2}{25\pi^2\hbar^4}(2\pi)^{-3}\mathrm{A}_{\mathrm{T}}^2\mathrm{A}_{\alpha}^2\mathrm{V}^2(1-g)^2\,|\,\mathrm{I}_1+\mathrm{I}_2-2\mathrm{I}_3\,|^2, \qquad \ldots \quad (9)$$

with I_1, I_2, I_3 as given in (6) but with ϕ_{F} replaced by

$$\phi_{\mathrm{F}}=\exp(i\mathbf{K}\,.\,\mathbf{r})\exp(i\mathbf{k}'\,.\,\mathbf{R})\phi_{\mathrm{T}}/\mathrm{A}_{\mathrm{T}}.$$

I_3 can be shown to be a delta function, which we can omit. I is a function of the momentum of the scattered neutron and of the momentum of the remaining triton, for example, related by the energy conservation condition. Then if we wish to get, for instance, the angular and energy distribution of the triton we have to integrate over the two parameters describing the scattered neutron.

We shall carry out this integration in the laboratory system, in which the momentum of the triton will be denoted by $\hbar\mathbf{k}_0$.

Expressing $\mathbf{k}'$ as a function of $\mathbf{k}_0$, $\mathbf{k}$ and $\mathbf{K}$, we get for the differential cross section

$$d\sigma = \mathrm{I}\frac{\mathrm{K}}{k}d\mathbf{k}_0\,d\Omega_{\mathrm{K}},$$

$d\Omega_{\mathrm{K}}$ can be replaced by $d\mathbf{K}\,\mathrm{K}^{-2}\delta(\mathbf{K}-\gamma)$ where γ is such that for $\mathrm{K}=\gamma$ the energy conservation relation is satisfied. Furthermore we have

$$\delta(\mathrm{K}-\gamma)=2\mathrm{K}\delta(\mathrm{K}^2-\gamma^2)$$

and for the energy of the system,

$$\mathrm{E}=\frac{5\hbar^2}{8\mathrm{M}}\mathrm{K}^2+\text{ terms which are independent of }\mathbf{K}.$$

Then

$$d\sigma = \mathrm{I}\frac{\mathrm{K}}{k}d\mathbf{k}_0\,d\mathbf{K}\,\mathrm{K}^{-2}2\mathrm{K}\delta\left[\frac{8\mathrm{M}}{5\hbar^2}(\mathrm{E}-\epsilon_0)\right],$$

where ϵ_0 is the initial energy.

Putting $\mathbf{W}=\mathbf{K}+\frac{1}{2}(\mathbf{k}\ -\frac{3}{4}\mathbf{k})$, the modulus of $\mathbf{W}$ being a constant on account of the energy conservation relation, $d\sigma$ can be written

$$d\sigma = \frac{5}{8}\mathrm{I}\frac{\mathrm{W}}{k}d\mathbf{k}_0\,d\Omega_{\mathrm{W}}\frac{8\mathrm{M}}{5\hbar^2}d\mathrm{E}\delta\left[\frac{8\mathrm{M}}{5\hbar^2}(\mathrm{E}-\epsilon_0)\right]$$

and integrating over E,

$$d\sigma = \frac{5}{8}\mathrm{I}\frac{\mathrm{W}}{k}d\mathbf{k}_0\,d\Omega_{\mathrm{W}}.$$

I is of the form $\sum_{i=1}^{3} f_i(\mathbf{k},\mathbf{k}_0)\exp[\mathbf{W}\,.\,\mathbf{V}_i(\mathbf{k},\mathbf{k}_0)]$, so that the integration over $d\Omega_{\mathrm{W}}$ is direct, and finally we get

$$\left(\frac{d\sigma}{d\Omega}\right)_0 = \frac{\mathrm{W}}{k}\mathrm{k}_0\sum_{i=1}^{3}\Gamma_i\frac{\sinh \mathrm{WV}_i}{\mathrm{WV}_i}\exp(\mathbf{A}_i\,.\,\mathbf{B}_i)\,d\mathrm{E}_0,$$

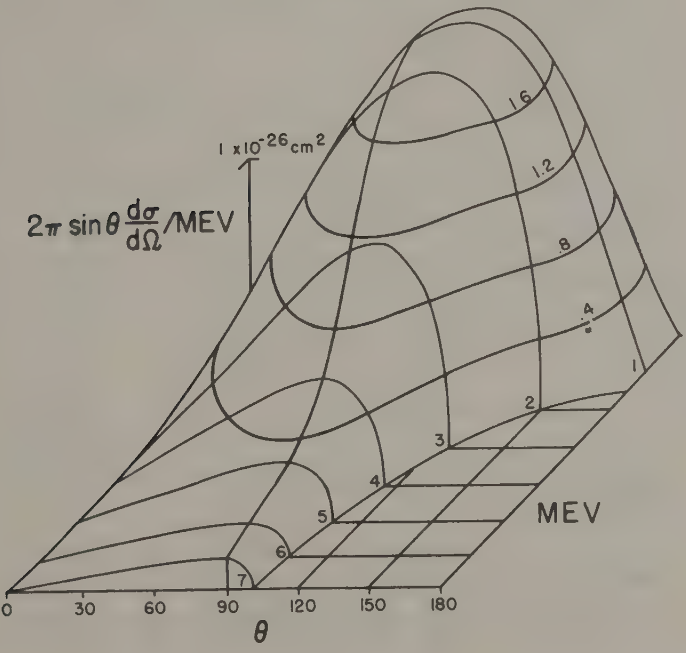
where $\mathbf{V}_i$, $\mathbf{A}_i$ and $\mathbf{B}_i$ are vector linear functions of $\mathbf{k}$ and $\mathbf{k}_0$ with coefficients obtained from the expressions for I_1 and I_2, Γ_i are coefficients and $\mathbf{E}_0$ the energy of the triton.

(2). *Results.*

The result of the numerical calculation is shown in fig. 2 which gives $2\pi \sin \theta_0 (d\sigma/d\Omega)_0$/MeV., the angular and energy distribution of the remaining triton in the laboratory system when a proton has been stripped off by the incoming neutron.

It is seen that these tritons are of low energy, 2 MeV. on the average, and that they are distributed almost isotropically ; this is seen more easily in fig. 3 where $(d\sigma/d\Omega_0)$/MeV. is plotted for four various energies of the triton.

Fig. 2.



A graphical estimation of the total cross section for stripping of a proton gives

$$\sigma_P = (0{\cdot}63 \pm 0{\cdot}10) \times 10^{-25} \text{ cm.}^2.$$

(b) *Stripping Scattering of a Neutron.*

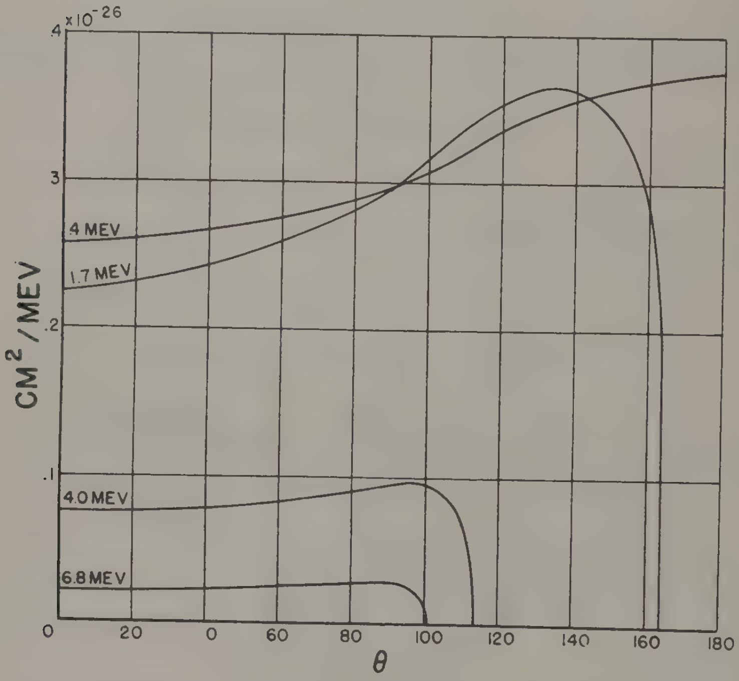
We have to perform the same calculations with ψ_F in (8) replaced by

$$P^{\sigma}_{13} P^{\sigma}_{24} P^{\chi}_{13} \psi_F.$$

But instead of two integrals as in (9) we are left with six of them, since the symmetry of this case is less than in the case of the stripping of a proton. As the subsequent integration over $d\Omega_W$ would be rather tedious we shall estimate the total cross section by the following argument : with the chosen potential between nucleons the neutron–neutron interaction takes non-zero values only in singlet states ; it can then be estimated to be $\frac{\frac{1}{4}}{\frac{1}{4}+\frac{3}{4}}\cdot\left(\frac{V_S}{V_T}\right)^2=$ one tenth of the neutron–proton interaction (Chew 1948). Assuming that the stripping of a nucleon off an alpha-particle is a process of the first order, we can say that the cross section for stripping of a neutron would be about

$$\sigma_N=0{\cdot}06\times10^{-25}\text{ cm.}^2.$$

Fig. 3.



3. Comparison with Experiments.

There are no experiments on 90 MeV. neutron–He scattering ; therefore we have to compare our calculations with experiments on other nuclei, especially the nuclei of carbon and oxygen, which may be considered as built of α-particles.

(A). *Total Cross Section.*

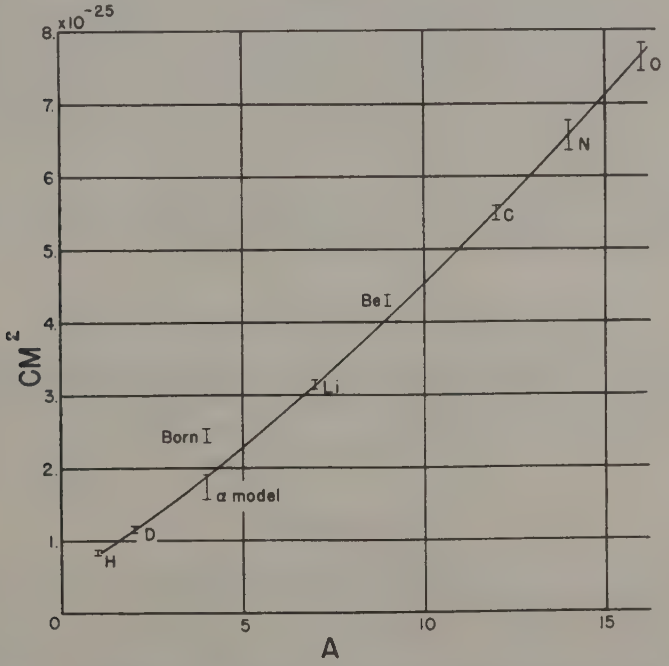
From the previous analysis we found for the total cross section for 90 MeV. neutron–He scattering the value

$$\sigma_\alpha=(2{\cdot}45\pm0{\cdot}10)\times10^{-25}\text{ cm.}^2$$

if we neglect the processes (iv.)–(vi.). In fact we think the cross sections for these processes to be very small, through the following estimate: let P be the probability for the incoming neutron to strip off a proton and N for stripping of a neutron, and S the probability for the incoming neutron to stick to a stripped proton. These three numbers can be estimated from the previous calculations. Then the cross section for process (v.) for instance. could be estimated to be 1 per cent the elastic cross section $\sigma_{(V)}=PN\sigma_{el}$.

Experimental data about the total cross section can be obtained from the measurements performed on the nuclei H, D, Li, Be, C, N, O . . . (Cook *et al.* 1949). If we draw a curve plotting the experimental values of σ as a function of the atomic mass, we see that the value for He would be about 2·0 10^{-25} cm.^{2} (fig. 4).

Fig. 4.



In order to check this estimation we shall use the following argument: Let us consider the carbon and oxygen nuclei as a triangle and a tetrahedron of side *s*, which we can reasonably assume to be the same in C and O, (Wheeler 1937) with weakly bound alpha particles at the corners; then we can, with the Born approximation, calculate their cross sections from that of He as a function of *s*; conversely from the experimental values of the cross sections of C and O, we can calculate *s* and σ_α. We find

$$\sigma_c=3(1+2y)\sigma_\alpha, \qquad \sigma_0=4(1+3y)\sigma_\alpha,$$

with

$$y = \sin ks/ks$$

and, comparing these values of σ_c, σ_0 with the experimental values, we get

$$s = (1 \cdot 79 \pm 0 \cdot 08) \times 10^{-13} \text{ cm.}, \qquad \sigma_\alpha = (1 \cdot 70 \pm 0 \cdot 17) \times 10^{-25} \text{ cm.}^2.$$

This value for σ_α can be said to be in good agreement with the value we calculated, in view of the fact that our calculations are crude for the following reasons: (i.) use of Born approximation, (ii.) use of the plane wave approximation as made in (8), (iii.) use of gaussian functions for the radial part of the potential and for the wave functions, especially for the deuteron one (Chew 1948).

Nevertheless if one makes a better calculation for neutron–He scattering, the He cross section as deduced from these for the carbon and oxygen should be considered more carefully:

(i.) We have not taken into account, using the Born approximation, the influence of double scattering. A measure of this effect is given by the amplitude of the wave, scattered by a single α-particle, at the distance s from it; this we can write approximately from our previous calculations

$$4 \exp(-4\theta^2).$$

The average value of this is 0·25.

Then the correction will be proportional to the number of ways in which a pair of alpha particles can be chosen in C and O nuclei, *i. e.* 3 and 6. We can write

$$\sigma_c' = 3(1+2y')\sigma_\alpha' + 3\Sigma, \qquad \sigma_0' = 4(1+3y')\sigma_\alpha' + 6\Sigma,$$

where Σ is the correction term due to double scattering, and y' is defined in the same way as y but for a new value s' of s. As an estimate we can write

$$\Sigma \sim (0 \cdot 25)^2 \tfrac{1}{2} \sigma_\alpha,$$

hence

$$y' \sim y(1 - 0 \cdot 015).$$

For the value of the cross section of He the correction has a coefficient small compared with unity; thus we can say that σ_α is not influenced appreciably.

(ii.) We have not taken into account the binding energy of the alpha particles, which is 2·4 MeV. per bond. This binding will not affect our estimate of the elastic cross section of the alpha particles in C and O, nor the cross section for production of deuterons, as follows from inspection of their spectrum.

But the stripping cross section will be reduced: the nucleus of mass 3, which received on a small energy, will almost never be ejected; then both of the other nucleons have to receive an energy greater than $0 \cdot 6l$ MeV., with $l = 2$ or 3, in order that the stripping scattering should be possible. As an estimate we can take, for the differential cross section for the scattering of one of these two nucleons, the experimental Berkeley curve (Hadley *et al.* 1948) and find that the stripping cross section has to be multiplied by $(1-0{\cdot}035\,l)$.

As the stripping cross section is about one quarter of the total one, we get

$$\sigma'_c=3(1+2y')(1-\tfrac{1}{4}0{\cdot}035\times 2)\sigma'_\alpha,$$

$$\sigma'_0=4(1+3y')(1-\tfrac{1}{4}0{\cdot}035\times 3)\sigma'_\alpha.$$

hence

$$y'\sim y(1-0{\cdot}017), \qquad \sigma'_\alpha\sim\sigma_\alpha(1+0{\cdot}017).$$

This correction will increase σ_α.

(iii.) We have not taken into account the fact that in the expansion of the exact wave the probability of finding the nuclei C or O in the state described above is less than 1, as the alpha particles are always exchanging nucleons, decomposing, reforming, . . . (Wheeler 1937). The most probable state after the one we used should be that in which one of the alpha particles is decomposed in lighter elements ; the probability of such an event is proportional to the number of alpha particles in C and O, *i. e.* 3 and 4.

If one of the alpha particles is decomposed we can write

$$\sigma_c=[2(1+y)+\rho]\sigma_\alpha,$$

$$\sigma_0=[3(1+2y)+\rho]\sigma_\alpha,$$

where $\rho\sigma_\alpha$ is the cross section of the decomposed alpha particle, together with interference terms which we shall neglect as is suggested from the preceding analysis by the small value of the correction on y ; then we can assume that ρ is the same for C and O. Then

$$\sigma'_c=3[1+2y'+q(\rho-1-4y')]\sigma'_\alpha,$$

$$\sigma'_0=4[1+3y'+q(\rho-1-6y')]\sigma'_\alpha.$$

q being the probability that an alpha-particle is decomposed. q being small, we get

$$y'\sim y-qy'(1+\rho), \qquad \sigma'_\alpha\sim\sigma_\alpha[1+q(1-\rho)],$$

σ_α will be increased or decreased according as $\rho\lessgtr 1$.

We can evaluate ρ as follows : the general slope of the curve of fig. 4 shows that the cross section of a light nucleus is less than the sum of the cross sections of the nuclei into which it could be decomposed. Assuming this to be true also for the alpha particle, we can write

$$\sigma_{\alpha d}>\sigma_\alpha,$$

where $\sigma_{\alpha d}$ denotes the cross section of the decomposed α-particle. Furthermore, since the elements into which an alpha particle could be decomposed have atomic masses in the range 1 to 3, we can say that the binding energy of the nucleons of these elements per nucleon will be in the range 0 to 2·8 MeV., while in the alpha particle it has the value 7·05 MeV., which is much larger. Then we can consider these elements as almost free nucleons, that is to say that the most important part of their cross section will be the elastic one. Under these circumstances the reduction of the cross section of these elements due to their being bound in C or O will be very small; we can then write

$$\sigma_{\alpha d} \text{ in C or O} \sim \sigma_{\alpha d}.$$

Furthermore, we remark that the most probable decomposition of the α-particle should be the one into one nucleon+one nucleus of mass 3. Then, from the curve fig. 4, we get

$$\rho \sim 1{\cdot}2.$$

The cross section of He is then

$$\sigma'_{x} \sim \sigma_{\alpha}(1-0{\cdot}2q),$$

q being small we can neglect this correction and estimate, in the light of this discussion,

$$\sigma_{\alpha}=(1{\cdot}74\pm0{\cdot}17)\times10^{-25}\text{ cm.}^2.$$

(B) *Production of Deuterons.*

In paragraph 2((A)(*b*)) we found for the cross section for production of fast (∼70 MeV.) deuterons in neutron–alpha particle collisions the value

$$\sigma_{\mathrm{D},\alpha}=0{\cdot}16\ 10^{-25}\text{ cm.}^2.$$

Experiments have been performed mainly on carbon nuclei (Brueckner 1949, York 1949). York gives the estimate

$$\sigma_{\mathrm{D,C}}=\text{a few hundredths of a barn,}$$

which certainly includes deuterons of energy smaller than say 60 MeV. Bruecker's value for deuterons of energy larger than 60 MeV. can be obtained from his Table II. and fig. 1:

$$\sigma_{\mathrm{D,C}}=(0{\cdot}02\pm0{\cdot}01)\ 10^{-25}\text{ cm.}^2.$$

In order to compare our value for the Helium nucleus with these experimental values on the carbon nucleus it would be necessary to take into account the fact that, when formed by interaction of the incoming neutron with one of the alpha particles of carbon nucleus, the deuteron has some probability of being disintegrated by interaction with the two other alpha particles of the carbon nucleus. Nevertheless, we think it is safe to say that the cross section for production of fast deuterons on carbon should be of the order of two to three times $\sigma_{D,\alpha}$. In view of the crudeness of our calculations as mentioned above, we can say that our result is in agreement with both of these experimental values.

(C) *Percentage of Inelastic Scattering.*

Our calculations show that the ratio of the inelastic to the total cross section is 0·35. The experimental value measured on various nuclei, He excepted, lies in the range 0·4 to 0·5 (Knable *et al.* 1949). The low value for He could be explained by the absence of small binding energies in He.

This work was performed in the Wills Physics Laboratory, Bristol. I would like to thank Professor N. F. Mott for valuable discussions. Miss J. Littleton for a large part of the numerical calculations, R. H. Dalitz for friendly advice and criticisms and the British Council whose scholarship enabled me to spend one academic year in Bristol University.

I would like to thank too Professor H. A. Bethe for criticisms.

REFERENCES.

BRUECKNER *et al.*, 1949, *Phys. Rev.*, **75**, 1274.
CHEW, 1848, *Phys. Rev.*, **74**, 809.
COOK, *et al.*, 1949, *Phys. Rev.*, **75**, 7.
GERJUOY and SCHWINGER, 1942, *Phys. Rev.*, **61**, 138.
HADLEY, *et al.*, 1948, *Phys. Rev.*, **73**, 1114.
HOISINGTON, *et al.*, 1939, *Phys. Rev.*, **56**, 884.
KNABLE, *et al.*, 1949, *Phys. Rev.*, **75**, 1470.
MOTT and MASSEY, 1933, *Theory of Atomic Collisions*, p. 25.
WHEELER, 1937, *Phys. Rev.*, **52**, 1083, 1107.
YORK, 1949, *Phys. Rev.*, **75**, 1467.

XL. *Relaxation Methods applied to Problems of Plastic Flow.*—II.

By J. A. JACOBS, M.A., Ph.D., A.F.R.Ae.S.,
Royal Holloway College, University of London *.

[Received January 11, 1950.]

SUMMARY.

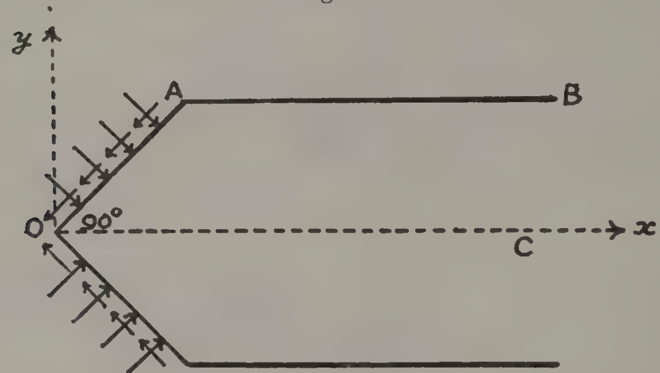
As in Part I., this paper considers the consequence of a particular hypothesis regarding plastic strain, viz. the Mises-Hencky criterion, and, by application of a numerical method due to Allen and Southwell, traces the effect of plastic straining. Two problems are considered here :

(i.) a sharp ended bar under specified edge tractions,

(ii.) the uniform compression of a block between two smooth rigid plates, which do not overlap it.

The general theory is not discussed, the reader being referred back to Part I.

Fig. 1.



1. STRESS DISTRIBUTION IN A SHARP-ENDED BAR UNDER SPECIFIED EDGE TRACTIONS.

Fig. 1 illustrates the sharp-ended bar subject to specified edge tractions, in which the normal and shearing forces over the nose are of equal magnitude. From symmetry it will be sufficient to consider one half the bar, BAOC. The bar was taken sufficiently long so that points sufficiently far from the nose are sensibly stress free. Thus the length of the nose AO is immaterial. As in Part I. (Jacobs 1950) dimensional factors are

---

* Communicated by the Author.

eliminated and in the subsequent analysis, the symbols stand for purely numerical quantities. Thus N_ν was taken equal to 960, and N_s equal to -960. (See fig. 2.)

The boundary conditions give, along OA,

$$N_\nu \cos\frac{\pi}{4} + N_s \sin\frac{\pi}{4} = X_\nu = \frac{\partial}{\partial s}\left(\frac{\partial\chi}{\partial y}\right),$$

i. e.

$$\frac{\partial}{\partial s}\left(\frac{\partial\chi}{\partial y}\right) = 0 \quad \text{or} \quad \frac{\partial\chi}{dy} = \text{constant}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

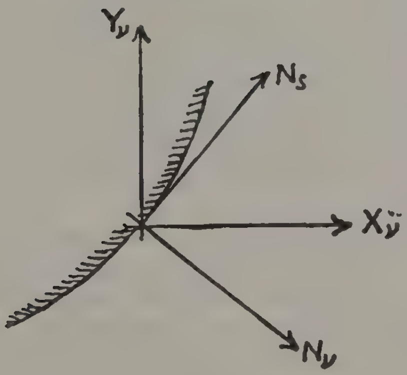
and

$$-N_\nu \sin\frac{\pi}{4} + N_s \cos\frac{\pi}{4} = Y_v = -\frac{\partial}{\partial s}\left(\frac{\partial\chi}{\partial x}\right),$$

or

$$960\sqrt{(2)} = \frac{\partial}{\partial s}\left(\frac{\partial\chi}{\partial x}\right)$$

Fig. 2.



which gives

$$960\sqrt{(2)}s + C = \frac{\partial\chi}{\partial x},$$

where C is a constant. Along OA, $s = x\sqrt{(2)}$, so that

$$1920x + C = \frac{\partial\chi}{\partial x}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Along AB, $N_\nu = 0 = N_s$. Hence

$$\frac{\partial\chi}{\partial y} \quad \text{and} \quad \frac{\partial\chi}{\partial x} \quad \text{are both constants.} \quad . \quad . \quad . \quad . \quad (3)$$

If we take $\frac{\partial\chi}{\partial y} = 0$ at O, then $\frac{\partial\chi}{\partial y} = 0$ along OA and AB, and if $\frac{\partial\chi}{\partial x} = 0$ at A, then from (2),

$$1920+C=0 \quad i.e. \quad C=-1920, \quad \text{since at A,} \quad x=y=1.$$

Hence along OA,

$$\frac{\partial \chi}{\partial x}=1920(x-1) \quad \text{and along} \quad \text{AB,} \frac{\partial \chi}{\partial x}=0. \quad . \quad . \quad . \quad (4)$$

Finally along OA, $\chi=960(x-1)^2+\text{constant}$.

Taking $\chi=0$ at A($x=1$), the constant is zero, so that

$$\chi=960(x-1)^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Since both $\frac{\partial \chi}{\partial x}$ and $\frac{\partial \chi}{\partial y}$ are zero along AB,

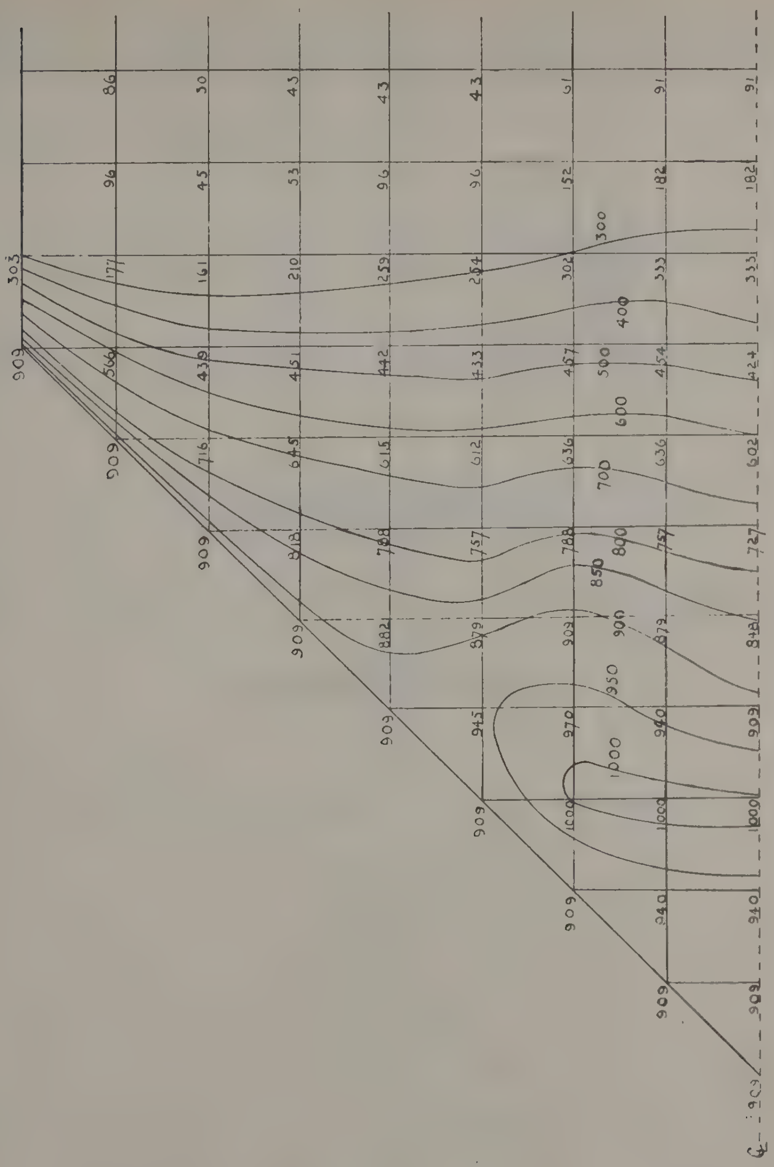
$$\chi=\text{constant along AB}$$
$$=\text{value of } \chi \text{ at A}=0.$$

The problem was first solved for the elastic case, the values of the principal stress difference being given in fig. 3. It was found that the same maximum values of the Mises-Hencky equivalent stress were attained at all mesh points long a line at right angles to the centre line, so that the result of plastic straining is evident at once from the contours of fig. 3. A plastic enclave starts on the centre line at H (fig. 4) and quickly extends outwards towards the faces of the bar. The growth is extremely rapid, and it would appear that once plastic flow has started in a region around H, it would very soon reach across the bar, isolating a small elastic core at the nose.

This problem was undertaken as a preliminary investigation into some of the problems that occur in sheet extrusion and sheet drawing, and which are discussed in detail by Hill (1948). One of the problems raised by Hill involves the reverse procedure to that considered here, viz. that of unloading from an initial plastic state. That problem, however, is not the same as loading an initially unstressed body of the same shape up to the final external stress system. In the practical problem raised by Hill, the material is unloading from an initial plastic state at the left hand end (the nose), and the material at the right hand end (at a great distance from the nose) is in a state of residual stress. It is these residual stresses which are of interest to the engineer. One must also consider the question as to whether the material immediately unloads or remains plastic for a time.

Assuming all the material to remain elastic during an unloading process, we can calculate the reversed stresses thereby induced, and superposing these on the stresses computed previously we have the stresses that will persist, unless plastic strain occurs in the process of unloading. It will do so, if in any region, these computed stresses violate the Mises-Hencky criterion, in which case adjustment will have to be made both in the stresses and in the strains.

Fig. 3.



Values of the principal stress difference for the sharp ended bar under specified edge tractions.

2. Uniform Compression of a Block Between Two Smooth Rigid Plates, which do not Overlap it.

Fig. 5 illustrates the block compressed between two smooth rigid plates which do not overlap it. The dimensions taken were

$$OD=DC=2CB.$$

The lengths of the plates were thus twice the block thickness. From symmetry it will be sufficient to consider the quadrant OABC, which is shown in fig. 6 re-orientated to compare with the problem of the notched

Fig. 4.

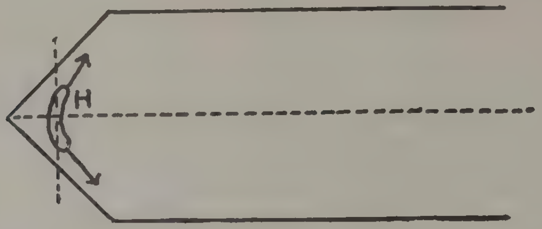


Fig. 5.

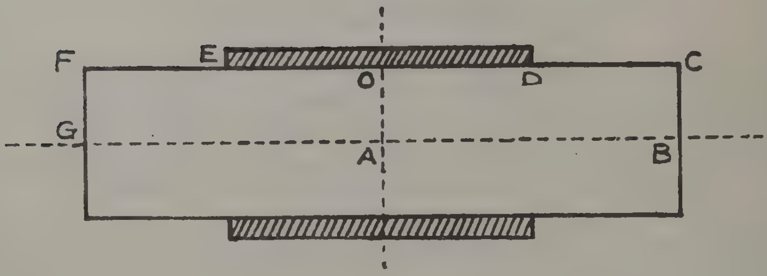
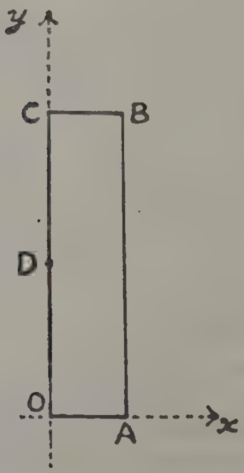


Fig. 6.



bar under tension discussed in Part I. (Jacobs 1950). This problem is, in fact, very similar to that of the notched bar, differing only in that the specimen is very much shorter in length (*i.e.* OA < AB, whereas before

OA > AB), and that a symmetry condition obtains at the right hand boundary AB, instead of a uniform tension there. As in the first two problems, dimensional factors are eliminated, and in the subsequent analysis, the symbols stand for purely numerical quantities.

The boundary conditions are

$$\left.\begin{array}{rl} \text{(i.)} & \text{BCD is stress free;} \\ \text{(ii.)} & \text{symmetry exists across OA and AB;} \\ \text{(iii.)} & \text{along OD, } Xy=0 \text{ (the plates are smooth) and} \\ & u=\text{const. (uniform compression).} \end{array}\right\} \quad . \quad . \quad (6)$$

From these last two conditions (iii.), we can deduce that, along OD,

$$\frac{\partial\chi}{\partial x}=0, \quad \text{and} \quad \frac{\partial^3\chi}{\partial x^3}=0.$$

Since BCD is stress free, both $\frac{\partial\chi}{\partial x}$ and $\frac{\partial\chi}{\partial y}$ are constant along it.

Owing to the symmetry condition, $\frac{\partial\chi}{\partial x}=0$ along AB.

Hence

$$\frac{\partial\chi}{\partial x}=0$$

along BC and CD. (7)

The constant value of $\frac{\partial\chi}{\partial y}$ along BC and CD was taken as 2000. Finally

$$\begin{aligned} \chi &= \int\left\{\frac{\partial\chi}{\partial x}\frac{dx}{ds}+\frac{\partial\chi}{\partial y}\frac{dy}{ds}\right\}ds \\ &= \int\left\{\frac{\partial\chi}{\partial y}\frac{dy}{ds}\right\}ds \quad \text{along} \quad \text{BCD} \\ &= \int 2000\,dy \\ &= 2000y+C. \end{aligned}$$

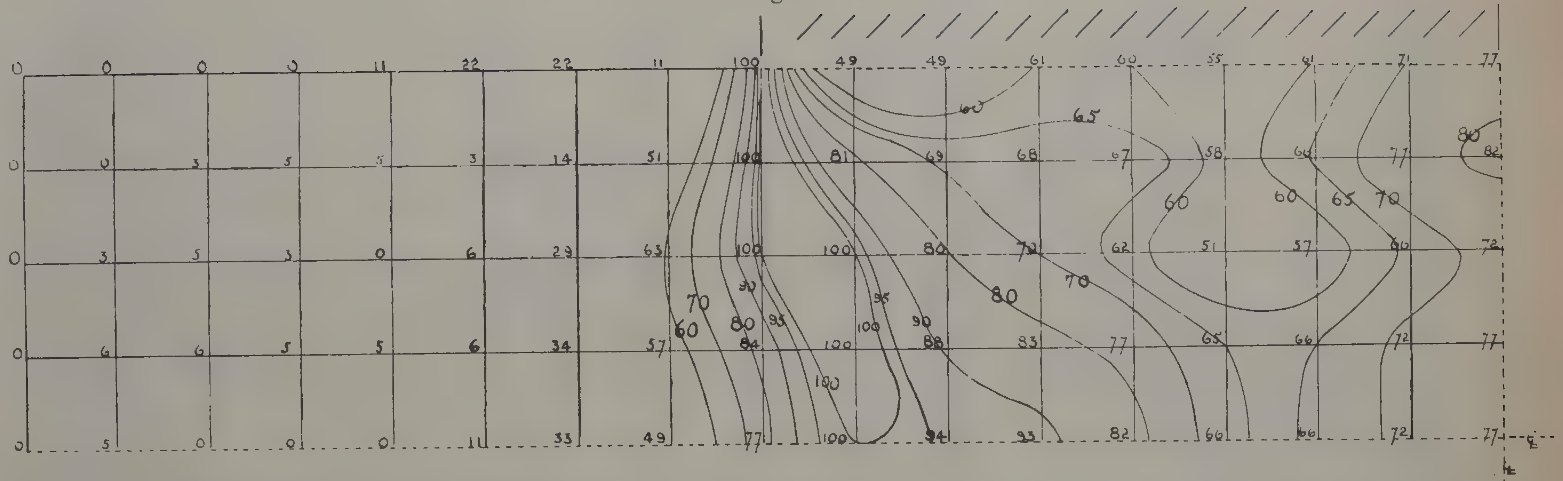
Taking $\chi=0$ at D ($y=\frac{1}{2}$), C = −1000, so that along BCD,

$$\chi=1000(2y-1). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Actually along BC ($y=1$), $\chi=1000$, and along CD, $\chi=1000\,(2y-1)$.

The problem was first solved for the elastic case; the contours of principal stress difference compare favouably with the isochromatics obtained by Frocht (1947) who examined the stresses produced by a concentrated load acting through a rigid flat die (the load being applied so that the deflections under the die were constant).

Fig. 7.



Values of the Mises-Hencky equivalent stress (taking 100 as its maximum value), when the plastic enclave has spread right across the block.

The elastic solution was modified by the methods already described and successive stages in the growth of the plastic enclave traced. Three distinct stages of plastic straining were obtained, and fig. 7 gives the values of the Mises-Hencky equivalent stress when the plastic enclave has spread until it extends right across the block. Fig. 8 summarizes these three stages, and illustrates the growth of the enclave as the compression is increased. In both these diagrams the upper left-hand quadrant is represented.

It was found that when the enclave had extended about 80 per cent of the way down the block, there was a strong tendency for it to spread towards the centre line OA as well as towards the other centre line GA. (Fig. 5.) It was not possible to obtain any records of this phenomenon owing to the extreme rapidity with which the enclave was spreading for quite small increases in the compressive load. This trend is illustrated, however, by the contours of values 90 and 80 of fig. 7 and was further confirmed by a preliminary investigation when all χ values were increased fourhold. As the enclave reached the centre-line GA, it spread out towards the centre-line OA, covering a distance approximately equal to $\frac{1}{2}$EO.

Fig. 8.

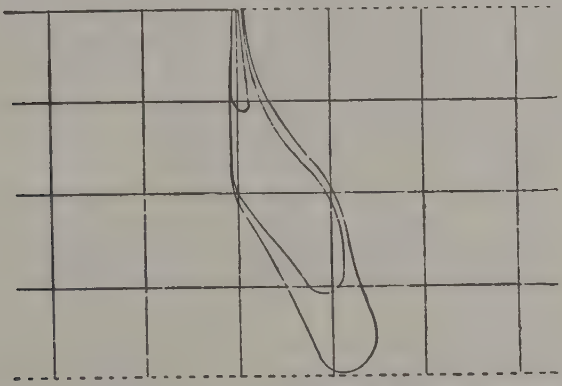


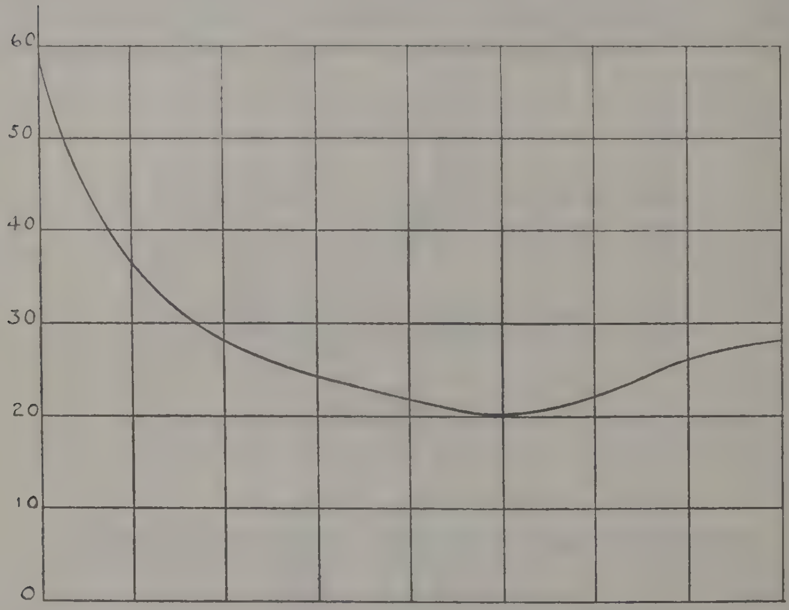
Diagram showing the progress of plastic straining (superposed plastic-elastic interfaces).

A feature of particular interest in this problem is the pressure distribution between the plates and the block, which is presented in fig. 9. The pressure falls off fairly rapidly from a maximum value at the ends of the plates, the decrease becoming less marked as we approach the centre of the plates, where it is interesting to note that the pressure intensity has risen again. The minimum pressure is about one third the maximum at the ends, whilst that at the centre is just under one half the maximum pressure. This pressure distribution was sensibly the same for the elastic case and all those of plastic straining, slight differences occurring only at the ends. The above distribution may be compared with the theoretical results of Sadowsky (1928), who showed that when a load is applied through a rigid flat die pressing on a semi-infinite plate so that the deflections under it are constant, the distribution of pressure under it is given by

$$p=\frac{\mathrm{P}}{\pi(a^2-y^2)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where P is the total applied load, and $2a$ the width of the plate. This gives $p=\mathrm{P}/\pi a$, a minimum at the centre ($y=0$), and makes p infinite at the ends ($y=\pm a$).

Fig. 9.



Pressure distribution between plate and block.

One of the reasons for the investigation of this problem was to compare the results obtained with the classical work of Prandtl on indentation. The problem of the indentation of a semi-infinite mass by a flat rigid-punch was first considered by Prandtl (1920), three years before Hencky (1923) published the general equations of the slip-line field. Hill (1949) has pointed out, however, that Prandtl does not seem to have fully appreciated the nature of the problem and the assumptions involved. "Prandtl did not realize the significance of the fact that the plastic boundary can only be determined by a calculation of the stresses in both plastic and elastic regions; nor was he aware of the condition determining the start of indentation, viz. that the plastic region should just include the slip lines from the centre O of the punch face to the free surface (since the slip lines are characteristics of the velocity equations)". Unfortunately, owing to the extreme rapidity with which the plastic enclave develops after it has reached the centre line of the block, it was impossible to determine with any great accuracy its further growth. It was not possible, therefore, to determine the plastic boundary when indentaton begins and so make direct comparison with Prandtl's work. The same difficulty, viz. that of determining the exact moment of indentation is also met with experimentally.

REFERENCES.

FROCHT, M. M., 1947, *Photo elasticity*, 2, 73 74 (John Wiley and Sons, Inc., New York).
HENCKY, H., 1923, *Z. Angew. Math. Mech.*, **3**, 241–251.
HILL, R., 1948, *Journ. Iron and Steel Inst.*, **158**, 177.
HILL, R., and TUPPER, S. J., 1948, *Ibid.*, **159**, 353.
HILL, R., 1949, *Quart. Journ. Mech. and Applied Math.*, 2, 40–52.
JACOBS, J. A., 1950, *Phil. Mag.*
PRANDTL, L., 1920, *Nachr. Ges. Wiss., Gottingen*, 74.
SADOWSKY, M. A., 1928, *Z. Angew Math. Mech.*, **8**, 107.

XLI. *Some Extensions in the use of Resistance Thermometry in the Study of Gaseous Explosions.*

By A. SMEETON LEAH, Ph.D., A.M.I.Mech.E.,
C. ROUNTHWAITE, B.Sc., and D. BRADLEY, Ph.D.,
University of Leeds*.

[Received January 25, 1950.]

[Plate XII.]

SUMMARY.

1. A modified technique for recording resistance wire temperatures is described.

2. By using plain and quartz-coated wires as the resistance elements in explosions a possible method of investigating the course of the reactions occurring within the flame reaction zone is suggested.

3. The corrections to be applied to estimate the true flame gas temperature from the quartz-coated wire records are evaluated.

4. A method of determining the rate of catalytic heating on the plain wires is outlined and the results from a typical explosion of a 90 per cent $CO+10$ per cent O_2 mixture at 0·4 atmosphere initial pressure are shown.

5. The thickness of the reaction zone in the early stages of the explosion is estimated to be 4 cm.

INTRODUCTION.

OF the known methods of measuring high temperatures in flame gases, particularly flame gases resulting from gaseous explosions, the resistance wire method has much to commend it. It is easy of application and can yield very reliable information when suitable precautions are taken and the necessary corrections applied. The fusing temperature of the wire used obviously sets a limit to the temperatures which can be recorded. David and co-workers(1) have consistently used this method for studying flame gas temperatures in closed-vessel explosions. The temperatures they have observed have generally been well below the flame temperatures measured by other methods for corresponding inflammable mixtures. Such discrepancies have led to a tendency in some quarters to discredit the method for general use(2). Yet it would appear to us that David(3) has explained qualitatively this seeming anomaly. Conditions in the flame gases resulting from explosions are very probably different from conditions above the inner zone in open flames. Still more convincing experimental proof of the substantial accuracy of the resistance method is afforded by experiments in which flame temperatures have been determined without introducing thermometers or other solid bodies into the flame gases(4). The mean translational temperature of the flame gases

---

* Communicated by the Authors.

was determined by correlating photographically determined flame volumes with pressure rise. It was found that the wire resistance thermometer temperatures agreed very closely with the pressure-volume temperatures, if anything, being slightly on the high side.

The usual argument for assuming the resistance method to yield incorrect temperatures appears to be that the inevitable radiation losses from the wires used are excessive. David and co-workers attempted to minimize this by the use of exceedingly fine wires of 0·0005 in. diameter and by this means reduced the error to small proportions. The correction to be applied to the observed wire temperature is amenable to fairly accurate computation and has been calculated for the experiments now to be described.

Of far greater consequence, however, is the possibility of errors caused by surface reactions on the wires due to the diffusion in towards the surface of active particles or uncombined gases. David and co-workers recognized that catalytic reactions might well occur on platinum wires and they therefore used special quartz-coated platinum wires on which no catalysis could occur. By this means the phenomenon of abnormal dissociation in carbon-monoxide explosions was discovered. The use of these two types of wires suggested to us a possible method of studying reactions occurring in the flame gases which has not, hitherto, received much attention.

If two wires, one completely inert to surface reactions and the other having a high catalytic surface activity to any particular reaction, could be used as resistance elements in an explosion, any difference in the temperatures attained by the wires would be a measure of the degree of catalytic heating occurring on the active wire. This principle forms the basis of the method developed by Kondratieva and Kondratiev(5) for detecting H atoms in flame gases. Platinum appears to be particularly effective in catalysing atomic and radical combinations, for example, Smith(6) found a reaction probability of unity for H atom recombination. Quartz-coated wires, on the other hand, are most probably relatively inactive, at least to reactions in carbon-monoxide explosions since they yield temperatures very close to the true translational temperatures of the flame gases. These two types of wires may, therefore, be taken to approximate to the ideal condition.

The technique of David and co-workers was developed in order to study the steady temperatures reached by the flame gases left behind the flame front. But it is certain that the greatest concentration of possible reactants, atoms or radicals or uncombined gases is likely to occur within the narrow zone immediately behind the flame front. In order to study this region a new technique has been developed utilizing a Cathode Ray Oscillograph for recording the rapidly changing wire temperatures. In addition a method of deducing the rate of catalytic heating of the platinum wires based upon dimensional analysis and requiring basic experimental data on the energy required to heat the wires electrically has been developed. These new techniques are described in the present paper. A later paper will show how certain interesting conclusions concerning the presence of atomic oxygen in carbon-monoxide explosions can be deduced from this type of experiment.

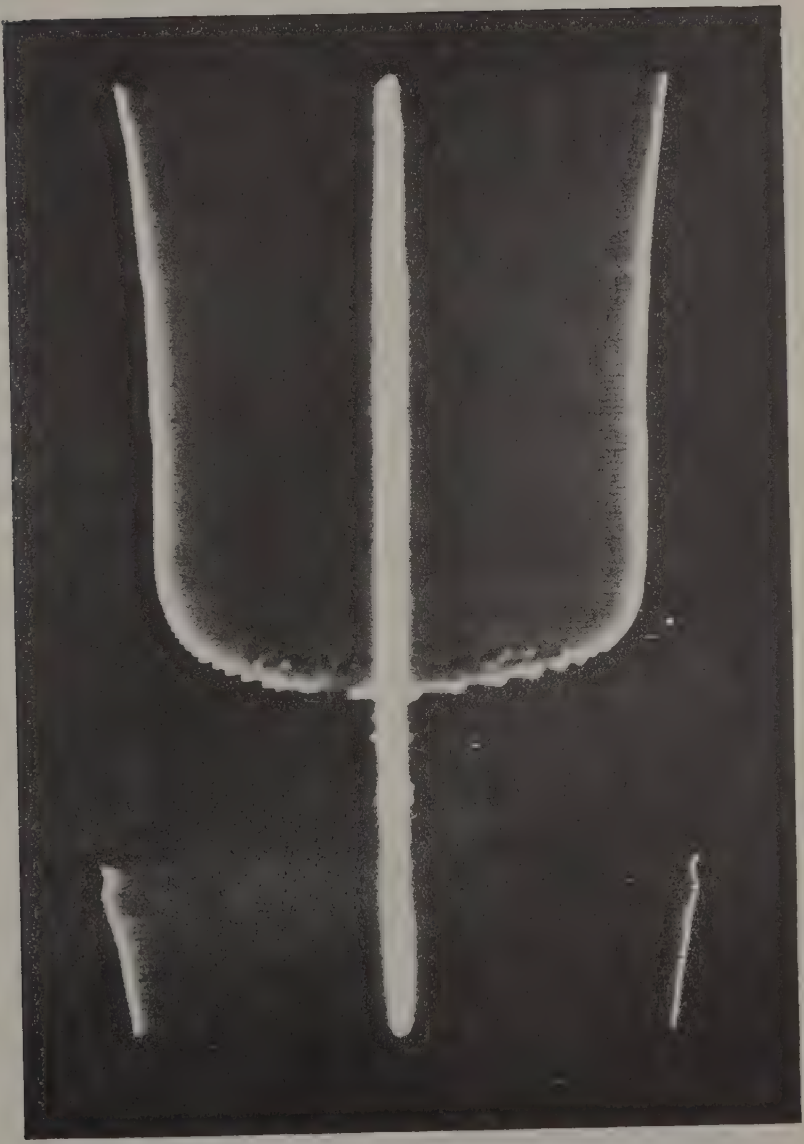
APPARATUS.

A spherical steel explosion vessel of 18 in. internal diameter and equipped with a centrally placed spark gap was used to carry out the explosions. Further details of the explosion vessel and of the earlier technique of recording wire temperatures may be obtained from a paper by David and Jordan⁽¹⁾. The present technique of recording the wire temperatures differed somewhat from that of David and Jordan. A resistance bridge was again used but the bridge was supplied with a high frequency alternating current of 5 milliamps at 1500 cycles per second from an oscillator and the recording was effected by the use of a Cathode Ray Oscillograph (Cosser Model 1035) equipped with a recording camera. The out-of-balance voltage caused by resistance changes of the wires during heating were recorded from the trace on the oscillograph screen as the envelope of the carrier wave curves. The photographic records were measured very accurately by the use of a Cooke, Troughton and Simms toolroom microscope adapted for the purpose.

The wires used were of two types namely, plain platinum-rhodium (10 per cent) wires of 0·0005 in. diameter and quartz-coated platinum-rhodium (10 per cent) wires of 0·0005 in. outside diameter of 0·0002 in. diameter metal core. The wires were electrically welded on to 22 gauge platinum-rhodium leads, in the case of the quartz-coated wires after dissolving off the quartz at the ends with hydrofluoric acid. A length of approximately 1 in. of wire was used in all cases and the wires were positioned $\frac{3}{4}$ in. distance from the spark gap.

The wires were directly calibrated by introducing them quickly into a tubular electric furnace and recording the deflections on the oscillograph screen. The wires were found to attain the furnace temperature in less than two seconds after immersion, the lag being due to the air currents set up by the motion of the wire holder. All the calibrations were therefore carried out at the end of a two-second period. Any prolongation of the immersion beyond a few seconds was found to result in a distinct change in the resistance characteristics of the plain wires due possibly to volatilization of some of the rhodium.

The inflammable mixtures of carbon-monoxide and oxygen were made up in a large tank of 11 cu. ft. capacity and were allowed to mix intimately before use. Similar precautions in drying the gases were taken as in earlier work⁽⁷⁾ and the gases used were also from the same source. The mixtures were ignited by means of a high tension electric spark provided by discharging a 3 microfarad condenser which had been previously charged to a sufficiently high voltage by a voltage generator. The apparatus used for determining the electrical energy required to heat up the wires was



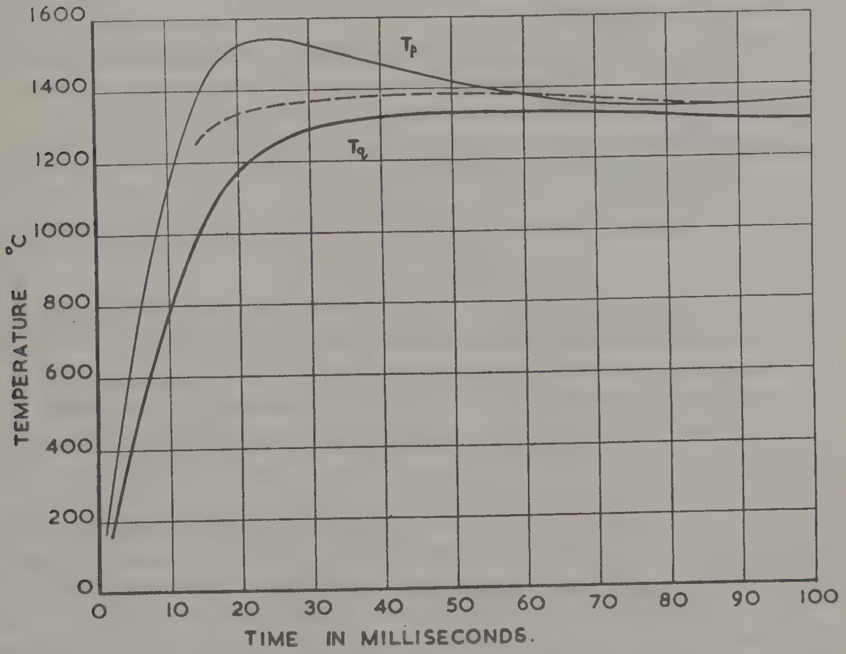
identical with that described by Davies[(8)]. These experiments were undertaken to provide basic data on which to estimate the rates of catalytic heating on the plain wires.

An enlarged reproduction of a typical oscillograph record of wire temperature taken with a quartz-coated wire during an explosion is shown in Pl. XII.

RESULTS OF EXPERIMENTS.

In a later paper it is hoped to give a comprehensive account of investigations with carbon-monoxide and oxygen explosions carried out at different initial pressures but for the present it would appear better to confine the attention to a single typical explosion for illustrative purposes only. A 90 per cent CO+10 per cent O_2 mixture explosion at an initial pressure of 0·4 atmosphere and an initial temperature of 15° C has been chosen for this purpose.

Fig. 1.



The plain and quartz covered wire temperature records are reproduced in fig. 1, the actual temperatures being plotted against the time from the commencement of the rise of temperature. The plain wire temperatures are given by the curve marked T_p and the quartz-covered wire temperatures by the curve T_q. The " hump " on the plain wire curve, T_p, will be apparent and indicates that considerable catalytic heating of the wire has occurred. In our view, these records may be taken to represent the changes of temperature of the wires themselves within very close limits. In order to estimate the true translational temperature of the flame gases it will be assumed, for the reasons previously mentioned, that no catalytic surface reactions of any kind occur on the quartz-covered wires. Nevertheless certain corrections must be applied in order to ascertain the true gas temperatures and these will now be considered in detail.

In a general case, let a wire having a temperature T at time t seconds be immersed in a gas maintained at a uniform temperature T_g in an enclosure whose wall temperature is maintained at temperature T_w. The wire will be either gaining or losing heat by conduction and convection from the gas and will be gaining or losing heat from the walls of the enclosure by radiation. In addition, let the wire be receiving heat by some other means (*e. g.* by electrical heating or catalytic surface heating) at the rate of S watts per sq. cm.

If H = the rate of heat loss by conduction and convection in watts per sq. cm. of wire surface.
R = the rate of heat loss by radiation in watts per sq. cm.
m = mass of wire in gm. per sq. cm. of surface.
k = mean specific heat of wire in watt-sec. per gm. per °C.

the equation of heat flow for the wire will be :—

$$mk\frac{dT}{dt}=S-R-H. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The function R is dependent only upon the temperature of the wire and that of the vessel walls. It is the same for a wire *in vacuo* as for a wire in a surrounding gas. The function H is dependent on the gas temperature and on the wire temperature. A method of determining these functions will be given shortly but first consider the curves of fig. 2 showing the functions R and H for a particular case, the function H being for a gas temperature $T_g=1300°$ C.

For quartz-coated wires it has been assumed that there is no catalytic heating, thus S=0. When the wire temperature has attained a steady value, T_s, the rate of heat accumulation $mk(dT/dt)=0$, thus R+H=0.

The wire temperature, T_s, will be given by the point A on the diagram where this condition is fulfilled, for there, the curve (H+R) cuts the temperature axis.

Alternatively, if T_s is known the gas temperature T_g may be found in the following manner. The function H is determined for various assumed values of T_g and the corresponding (H+R) curves are plotted. They are found to be very nearly parallel curves as will be seen from the broken line curves of fig. 2 which refer to gas temperatures of 1110° C. and 900° C. If the wire temperature were raised to equal the gas temperature T_g, the conduction and convection heat transfer would be zero, or H=0. The point B in fig. 2, therefore gives the gas temperature T_g which is also given by the point of intersection, D. of the curve R and the curve (H+R)*.

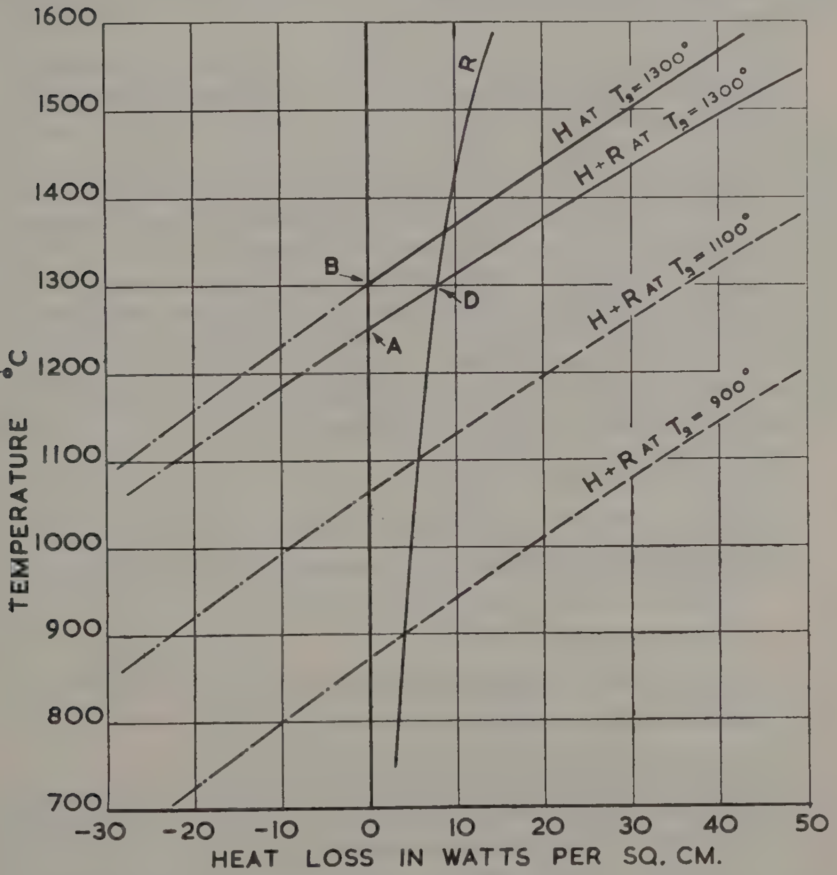
---

* This principle forms the basis of the heated wire method of determining flame temperatures used by Schmidt (9) and Kohn (10).

Thus with T_s known the point A is located and the (H+R) curve passing through this point is continued to intersect the curve R in D giving the gas temperature T_g. If the wire temperature is below the steady value the wire will be receiving heat at the rate :

$$mk\frac{dT}{dt}=-(H+R) \quad \text{(from equation (1))}.$$

Fig. 2.



From the record of the temperature changes of the wire, dT/dt may be determined at any time t as well as the wire temperature T. The gas temperature may again be inferred from the curves of fig. 2 by marking a point with co-ordinates $-mk(dT/dt)$ and T and following the (H+R) curve passing through this point until it cuts the curve R in D giving the gas temperature T_g.

It will be evident, therefore, that providing the functions H and R can be determined the quartz-coated wire temperatures may be corrected to yield the flame gas temperatures.

The radiation loss R was determined by the use of the Stefan-Boltzman law using values for the emissivity of quartz surfaces given by Pirani(11) up to fairly high temperatures. Doubtless some of the radiation from the platinum-rhodium core will be transmitted through the quartz coating resulting in a effective emissivity somewhere between those for the two materials. But since the two emissivities are not appreciably different at high temperature little error will result from using the value for quartz. For the case of platinum-rhodium wires, which will be dealt with later, values for the emissivity were Hottel's values taken from Brown and Marco(12).

The determination of the rate of conductive and convective loss values, H, is a matter of some considerable difficulty. The flame gases in an explosion are in motion due to the adiabatic compression resulting from the rising pressure. This motion will set up a forced convection loss from the wires in addition to the natural convection loss. Fortunately, in the very early stages of an explosion, whilst the flame diameter is small in relation to the diameter of the explosion vessel, the pressure rise is so small that the gas velocities due to compression are small enough to be neglected in relation to the natural convection velocities. With this simplification it is possible to predict the conduction and natural convection losses from electrically heated wire experiments by means of dimensional analysis.

The rate of heat loss by conduction and convection, H per sq. cm. of wire surface will be dependent on the following factors :—

D, the diameter of the wire.

T_g, the temperature of the surrounding gas.

T_d, the temperature difference between the wire and the surrounding gas.

P, the pressure of the gas.

K, the thermal conductivity of the gas.

μ, the viscosity of the gas.

C_p, the specific heat of the gas at constant pressure.

g, the acceleration due to gravity (to allow for the buoyancy of the gas undergoing natural convection).

Thus,

$$H=\phi[D, T_g, T_d, P, K, \mu, C_p, g].$$

This relation must be dimensionally homogeneous and can be expressed by a relation involving five dimensionless quantities, the most convenient for the present investigation being :

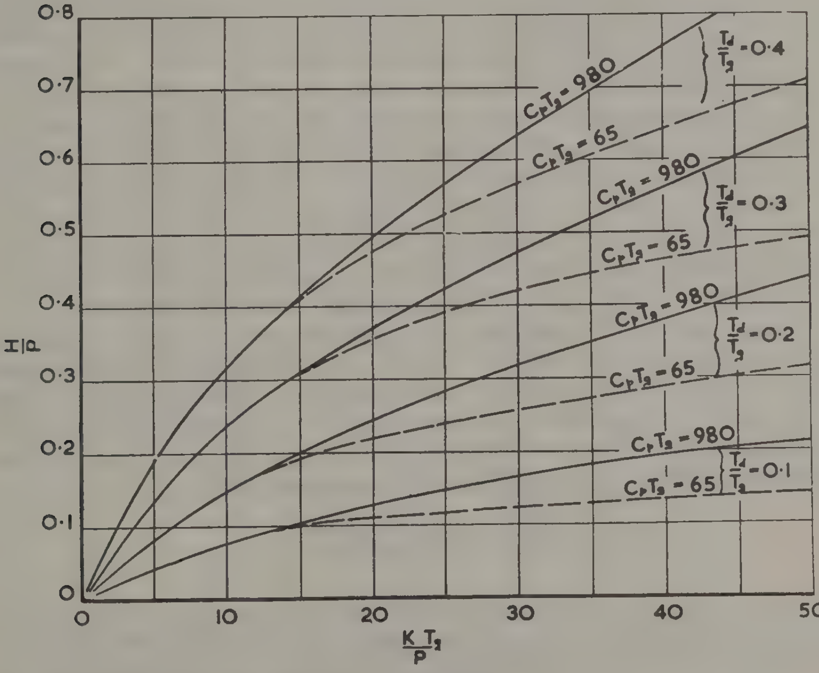
$$\frac{H}{D^{1/2}Pg^{1/2}}=\psi\left[\frac{T_d}{T_g}, \frac{KT_g}{D^{3/2}Pg^{1/2}}, \frac{C_p\mu}{K}, \frac{C_pT_g}{Dg}\right]. \quad . \quad . \quad . \quad (2)$$

The quantity $C_p\mu/K$, Prandtl's Number, is very nearly constant for all diatomic and triatomic gases and in addition, the wire diameter D is constant, thus equation (2) may be reduced to :

$$\frac{H}{P} = \psi'\left[\frac{T_d}{T_g}, \frac{KT_g}{P}, C_pT_g\right]. \qquad (3)$$

The relationship between the quantities in equation (3) was determined by measuring the power required to heat, electrically, in various quiescent gases at different pressures from 2·5 to 250 mm. Hg Abs., samples of the wires used in the explosion experiments. The wires were heated from atmospheric temperature up to about 400° C, and since the radiation losses are but a very small fraction of the conduction and natural convection losses at these low temperatures, the power input represents very closely the conduction and convection rate of loss. The results of these experiments are shown in fig. 3, values of H/P being plotted against KT_g/P for different values of the ratio T_d/T_g and for two gases, namely air and hydrogen, having widely different values of C_pT_g. It will be seen that the value of C_pT_g has no effect at the lower values of KT_g/P and only a slight effect at the higher values. For this reason little error will result in performing a linear interpolation between the extreme values of C_pT_g for intermediate values. It may be mentioned that when the heating

Fig. 3.



experiments were conducted with carbon-dioxide the resulting curves agreed identically with those for air in fig. 3 as was expected since the heat capacities of CO_2 and air per unit mass are very nearly the same.

In order to estimate the conduction and natural convection loss from a wire heated in any gas, the values of KT_g/P and C_pT_g are computed for the known temperature and pressure conditions of the gas. From the curves of fig. 3, the value of H/P is found for different values of the ratio T_d/T_g and the corresponding values of H and T_d yield the heat loss curve, H, shown in fig. 2 from which the (H+R) curve may be constructed. The values of H will only be obtained for positive values of T_d, (*i. e.* when the wire is above the gas temperature). Fortunately the H curves are very nearly straight lines and may safely be extrapolated to negative values of T_d, giving negative heat loss values. Such extropolated values are shown in fig. 2 by the chain-dotted part of the H curve. The corresponding portions of the (H+R) curves are also shown chain dotted.

The quartz coated wire temperatures of fig. 1 have been corrected for radiation and heat capacity in the manner described and the corrected curves are shown by the broken line in fig. 1. These corrected values will represent the true gas temperature providing no surface heating has occurred due to catalytic reactions, as we have assumed to be the case. Should it happen that catalytic reactions do occur, then the broken line curve of fig. 1 will give over-estimates of the true gas temperature.

Turning now to the plain wire temperature record of fig. 1, here it is apparent that catalytic surface reactions are occurring causing the wire temperature to exceed the gas temperature. At any instant the wire temperature $T=T_p$ is known, the gas temperature T_g may be found from the corrected quartz-coated wire temperatures and the rate of rise of wire temperature dT/dt is known. Thus from equation (1):

$$S=H+R+mk\frac{dT}{dt}.$$

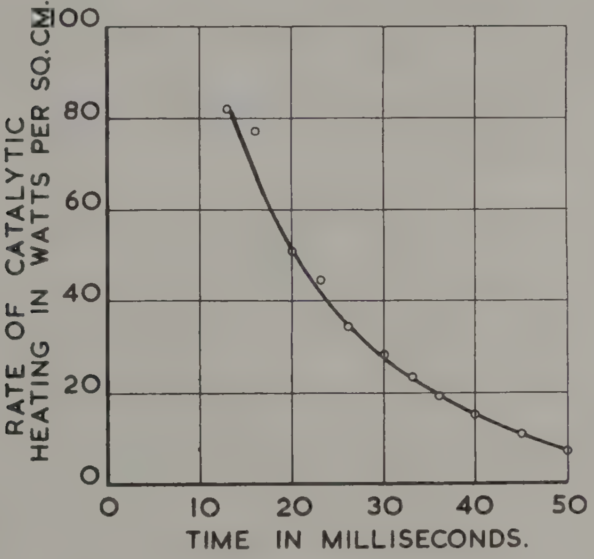
Knowing the gas temperature T_g, the function H may be determined and the curve of H+R constructed as shown in fig. 2. The value of H+R at the wire temperature T_p can be read off this curve and the rate of catalytic heat reception can then be directly computed.

Values of the rate of catalytic heating for the explosion under consideration are shown in fig. 4 plotted against the time from the commencement of the rise of temperature. As would be expected from the temperature curves of fig. 1, the rate of heating diminishes to zero after about 50 milliseconds from which it may be concluded that at this time after the passage of the flame front the reacting species have disappeared from the flame gases.

No attempt has been made in this work to measure the flame velocities by photographic, or other means. An estimate of the mean flame velocity can be found, however, from the time taken to reach maximum pressure after the passage of the igniting spark. The mean flame speed appears to be of the order of 80 cm. per sec. If the reaction zone in the advancing flame front is defined as that zone in which species exist which set up the catalytic surface reactions on the plain wires, its depth must be 0·05×80=4 cm. That is to say, the flame front will be 4 cm. beyond the wire when catalytic activity on the wire surfaces ceases. This is far and away above the thicknesses calculated by Lewis and von Elbe(13) for ozone explosions but is not surprising in view of the much slower burning gas we have used, quite apart from the very different chain mechanism by which the flame is propagated.

Further discussion will be deferred and given in a subsequent paper.

Fig. 4.



REFERENCES.

(1) David and Jordan, 1934 a, *Phil. Mag.*, **17**, 172 ; 1934 b, *Ibid.*, **18**, 228.
David, Leah, and Pugh, 1941, *Phil. Mag.*, **31**, 156.
(2) *Cf.* Lewis and von Elbe, 1935, *Engineering*, **139**, 168.
Ribaud, Laurie, and Gaudry, 1939, *J. Inst. Fuel*, **12**, 18.
(3) David, 1937, *Phil. Mag.*, **22**, 251.
David, Leah, and Pugh, 1937, *loc. cit.*
(4) Leah, 1943, *Phil. Mag.*, **34**, 795 ; *Ibid.*, 1947, **38**, 657.
(5) Kondratieva and Kondratiev, 1946, *Acta. Physicochimica, U.R.S.S.*, **21**, 1.
(6) Smith, 1943, *J. Chem. Phys.*, **2**, 110.
(7) Leah, 1947, *Phil. Mag.*, **38**, 657.
(8) Davies, 1934, *Phil. Mag.*, **17**, 233.
(9) Schmidt, 1909, *Ann. Physik*, **29**, 1027.
(10) Kohn, 1914, *Ann. Physik*, **44**, 749.
(11) Pirani, 1939, *J. Scientific Instruments*, **16**, 377.
(12) Brown and Marco, 1942, *Introduction to Heat Transfer*, McGraw-Hill.
(13) Lewis and von Elbe, 1938, *Combustion, Flames and Explosions of Gases*, Cambridge.

XLII. *Atomic Oxygen in Carbon-Monoxide Explosions.*

By A. Smeeton Leah, Ph.D., A.M.I.Mech.E., C. Rounthwaite, B.Sc.
and D. Bradley, Ph.D.,
University of Leeds †.

[Received March 2, 1950.]

Summary.

Catalytic heating on plain platinum wires has been studied by a method described in a previous paper and has been ascribed to the presence of high concentrations of atomic oxygen within the reaction zone of 10 per cent O_2+90 per cent CO gaseous explosions carried out in a large spherical explosion vessel. The atomic oxygen quickly disappears in the flame gases according to an overall reaction process which gives an effective first order reaction constant. The atomic oxygen content of the flame gases is also shown to account for only a small part of the total " missing " energy in the explosions.

Contemporary theories of flame propagation in gaseous explosions recognize the important part played by the diffusion of active particles such as atoms or radicals into the unburnt gases ahead of the advancing flame front. These active particles being capable of initiating chain branching reactions in the unburnt gases multiply their concentration to a value far in excess of the equilibrium value as was shown by Jost and Müffling (1938) and more recently by Tanford (1947). Making certain simplifying assumptions Lewis and von Elbe (1934, 1938) were able to evolve a theoretical treatment allowing for the effects of diffusion which enabled them to calculate the flame speeds in ozone explosions with remarkable accuracy. Their analysis showed that the concentration of atomic oxygen in the reaction zone rises from the equilibrium value in the unburnt gases ahead of the flame front, passes through a maximum and reverts to the corresponding equilibrium value in the burnt gases behind the flame front when all reactions are completed. It is not yet possible to carry out analyses similar to that of Lewis and von Elbe for other gaseous inflammable mixtures as insufficient is known of the kinetics of the chain reactions involved in the combustion process. There is no doubt, however, that whatever inflammable mixture is used, the concentrations of the particular active species taking part in the reaction will pass through well defined maxima somewhere in the reaction zone. There is certainly some doubt about whether the concentrations of the active species rapidly attain the normal equilibrium values in the burnt gases left behind the flame front since electronically excited molecules, and possibly vibrationally excited molecules, remain in these gases for a long time, as evidenced by the long-sustained afterglow.

---

† Communicated by the Authors.

In the case of carbon-monoxide, oxygen explosions, although a theoretical analysis cannot yet be undertaken, it has become possible in certain cases to determine experimentally the concentration of atomic oxygen in the reaction zone by utilizing the technique of wire resistance thermometry described in a previous paper (Leah, Rounthwaite and Bradley, 1950), by which the rate of catalytic heating of platinum wires could be determined. Consider for a moment a general case; if it were known that a particular active species would react catalytically on one type of resistance wire element with some other reactant present in large excess, whereas with some second type of resistance wire element no surface reactions at all could be sustained, then providing the reaction probability on the surface of the first type of wire were known, the observed rate of catalytic heating should be a direct measure of the rate of diffusion to the wire of the particular active species involved. From this an estimate could be made of the concentration of the active species in the surrounding gases. This principle has been utilized by Kondratieva and Kondratiev (1946) in their method of measuring the concentration of atomic hydrogen in the slow reaction flames of hydrogen–oxygen mixtures. They used two similar thermocouples, one coated with $ZnOCr_2O_3$ which has a high reaction probability towards hydrogen atom recombination and the other coated with KCl, which has a very low reaction probability, as was shown by Taylor and Lavin (1930) and W. V. Smith (1943).

In explosions of dry, hydrogen free, $CO–O_2$ mixtures it would appear that the only active species which are likely to be present in the flame gases are O atoms, excited oxygen molecules O_2^* and excited carbon-dioxide molecules CO_2^*. Gaydon (1946) has demonstrated the presence of O atoms in $CO–O_2$ open flames by means of the nitric oxide test. Similarly Kondratjew (1937) has shown that low pressure flames of $CO–O_2$ mixtures contain a large concentration of CO_2^*. The presence of these active species in the explosion flame gases is, therefore, highly probable. Laidler (1949) has shown recently that O_2^* is also present in flame gases resulting from the combustion of $CO–O_2$ mixtures containing excess O_2 but if the mixtures contain a large excess of CO then it seems unlikely that O_2^* will be present in large amounts during explosions. In fact it would seem that with these mixtures the most probable active species will be O atoms and CO_2^*.

In any case, since the mixtures contain a large excess of CO, it is certain that any O atoms or normal O_2 molecules (or excited O_2^* molecules if any are present) would react catalytically with the CO on plain platinum wires. The work of Langmuir (1921) indicates that the reaction probability would be very high on platinum and indeed Smith (1948) shows that platinum has a reaction probability of unity towards certain atomic and radical combinations. In addition, of course, there will be the possibility of recombination of O atoms taking place on the surface together with heating effects due to the normalizing of CO_2^* on striking the wire. With quartz-coated wires, however, the catalytic surface reactions will be inhibited although the surface effects due to CO_2^* and recombination of O atoms may still take place in some degree. If the heating due to these effects were substantial then the quartz-coated wires should record temperatures well in excess of the true translational temperature of the flame gases, due allowance being made for radiation losses from the wires. The quartz-coated wires, however, have been shown to yield values of gas temperatures which are not greatly in error (Leah 1943) and it must be concluded that only a relatively small surface heating occurs with these wires. But in any case it is reasonable to suppose that similar surface effects will occur on the plain platinum wires and hence, the differences in the temperatures observed with the two types of wires should still be a measure of the catalytic heating due to O_2 and O atom combination with CO on the plain wires.

Although it will not be possible to say what proportion of the observed catalytic heating of the plain platinum wires is due to O atoms or to O_2 molecules it is possible to prove that the heating is not due to the combination of O_2 with CO alone. Analysis shows that if it is assumed that the high rate of catalytic heating observed in the early stages of the explosions is due to O_2 alone, the concentration of the O_2 in the high temperature flame gases would require to be a very large proportion of that in the unburnt mixture, a condition which is obviously absurd. In fact it is evident that the catalytic heating must be due, in very large measure, to the combination of O atoms with CO on the platinum wires. Since the mixtures chosen contain a large excess of CO, and since it will be taken that the reaction probability of platinum for the O atom—CO reaction is unity, the rate of catalytic heating will be governed by the rate at which the O atoms diffuse to the wire surface from the surrounding gas.

As it is the object of the present work to estimate the concentration of the atomic oxygen in the flame gases well removed from the influence of the wires, it will be essential before proceeding further, to form some idea of the rate of reaction causing the disappearance of O atoms in the gas phase. In this connection the catalytic heating curve given in the previous paper, and reproduced later in fig. 3 of the present work, gives an indication of the rate of disappearance of the active species causing the catalytic heating of the wires irrespective of the particular identity of the species concerned. The rate of catalytic heating will be proportional to the rate of diffusion of the reacting species to the wire surface, and this in turn may be taken, as a first approximation, to be proportional to the concentration of the species in the surrounding flame gases. On plotting the logarithm of the rate of catalytic heating against time, as in fig. 1, the experimental points are seen to lie very nearly on a straight line. This indicates that the rate of reaction of the active species is given by the expression

$$\frac{d[x]}{dt}=g[x],$$

where g is a constant, and $[x]$ the concentration of the species. If the species is wholly atomic oxygen, then it will be reasonable to infer that $d[\mathrm{O}]/dt=g[\mathrm{O}]$. An alternative argument leads to the same conclusion as follows:

The most probable reactions occurring in the reaction zone would appear to be those suggested by Semenov (1935), namely,

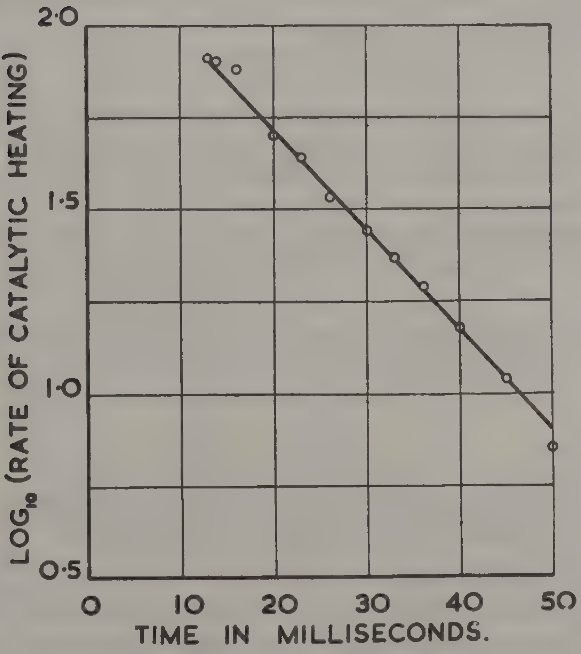
$$\mathrm{O+CO \rightarrow CO_2^*}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\mathrm{CO_2^*+O_2 \rightarrow CO_2+2O}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\mathrm{O+CO+M \rightarrow CO_2+M} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

(where M is any third body).

Fig. 1.



In addition there will be certain reactions involving $\mathrm{O_3}$ and probably the chain branching reaction suggested by Hinshelwood (1940),

$$\mathrm{O+CO+O_2 \rightarrow CO_2+2O},$$

although Semenov (1935) has pointed out certain objections to this reaction. The explosion flames would appear to be propagated chiefly by the branching chain reactions (1) and (2) whereas the disappearance of O atoms from the reaction zone would seem to be most probably accounted for in large measure by the chain breaking reaction (3).

When the flame front has passed beyond the wires, the temperature of the plain wires rises very rapidly corresponding to the period of most intense chain branching in the reaction zone. After this stage, which is the period when the O atoms are disappearing and with which period the present work is mainly concerned, the reactions then proceeding will be dominated by those of a chain-breaking character. The rate of decay of the O atoms will be given by

$$\frac{d[\mathrm{O}]}{dt}=f([\mathrm{O}],[\mathrm{CO}],[\mathrm{O}_2],[\mathrm{O}_3],[\mathrm{CO}_2],[\mathrm{CO}_2^*]). \quad . \quad . \quad . \quad (4)$$

It would appear that the greatest contribution to the function (4) is likely to be that due to reaction (3), and since CO is present in large concentration during the later stages in the reaction zone, then, at any one pressure and temperature, we may write as a first approximation

$$\frac{d[\mathrm{O}]}{dt}=g[\mathrm{O}]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Here g is very nearly a constant at any one pressure and temperature, and it will be shown later that, during the period of the explosions we are concerned with, the pressure and temperature do remain substantially constant and equation (5) will then apply.

The concentration of the atomic oxygen in the flame gases may be determined from the catalytic heating curve by means of the following analysis :

Let n'=the concentration of atomic oxygen in the flame gases away from the influence of the wires at time t.

Then

$$\frac{dn'}{dt}=gn'. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Let n=the concentration of atomic oxygen at a distance r from the centre of a plain platinum wire at time t,

a=radius of the wire,

D=diffusion coefficient for O atoms diffusing into the flame gases.

Neglecting the relatively small temperature gradient in the flame gases surrounding the wires the diffusion equation will be

$$\frac{dn}{dt}=\mathrm{D}\left\{\frac{\partial^2 n}{\partial r^2}+\frac{1}{r}\frac{\partial n}{\partial r}\right\}+gn. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Imagine the wire to be suddenly immersed in the gases at time $t=0$ and let it be assumed that $n=n'=n_0$ at $t=0$, an assumption which will be seen to be reasonably close in the later part of this paper.

Subsequently, since the reaction probability at the wire surface is unity $n=0$ at the wire surface at all times after $t=0$.

i. e. for $t>0$, $n=0$ at $r=a$.

In order to solve equation (7) let

$$n=n'(n''+1), \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (8)$$

where n' is defined by equation (6) and is independent of r and n'' is defined by

$$\frac{dn''}{dt}=\mathrm{D}\left\{\frac{\partial^2 n''}{\partial r^2}+\frac{1}{r}\frac{\partial n''}{\partial r}\right\}. \quad . \; . \; . \; . \; . \; . \; . \quad (9)$$

It can easily be shown that relations (8) and (9) satisfy the diffusion equation (7). The boundary conditions will be :

At $t=0$, $n''=0$ for $r>a$

and also for $t>0$, $n''=-1$ at $r=a$.

Carslaw and Jaeger (1947) show the solution of (9) with these boundary conditions to be

$$n''=-1-\frac{2}{\pi}\int_0^\infty \exp\,(-\mathrm{D}u^2t)\,\frac{\mathrm{J}_0(ur)\mathrm{Y}_0(ua-)\mathrm{Y}_0(ur)\mathrm{J}_0(ua)}{\mathrm{J}_0^2(ua)+\mathrm{Y}_0^2(ua)}\,\frac{du}{u}$$

and also that

$$\left(\frac{\partial n''}{\partial r}\right)_{r=a}=\frac{4}{a\pi^2}\int_0^\infty \exp\,(-\mathrm{D}z^2t/a^2)\,\frac{dz}{z[\mathrm{J}_0^2(z)+\mathrm{Y}_0^2(z)]}, \quad . \; . \quad (10)$$

where $z=au$.

For large values of $\mathrm{D}t/a^2=x$, Carslaw and Jaeger give an approximation for the right hand side of (10)

$$\left(\frac{\partial n''}{\partial r}\right)_{r=a}=\frac{4}{a\pi^2}\left[\frac{\pi^2}{2}\left\{\frac{1}{\log_e(4x)-2\gamma}-\frac{\gamma}{(\log_e(4x)-2\gamma)^2}\right\}+\text{etc.}\right], \quad (11)$$

where γ=Euler's constant=0·57722. From (8)

$$\left(\frac{\partial n}{\partial r}\right)_{r=a}=n'\left(\frac{\partial n''}{\partial r}\right)_{r=a}. \quad . \; . \; . \; . \; . \; . \; . \quad (12)$$

Substituting (11) in (12) and multiplying through by D

$$\mathrm{D}\left(\frac{\partial n}{\partial r}\right)_{r=a}=\frac{2\mathrm{D}n'}{a}\left[\frac{1}{\log_e(4x)-2\gamma}-\frac{\gamma}{(\log_e(4x)-2\gamma)^2}+\text{etc.}\right]. \quad (13)$$

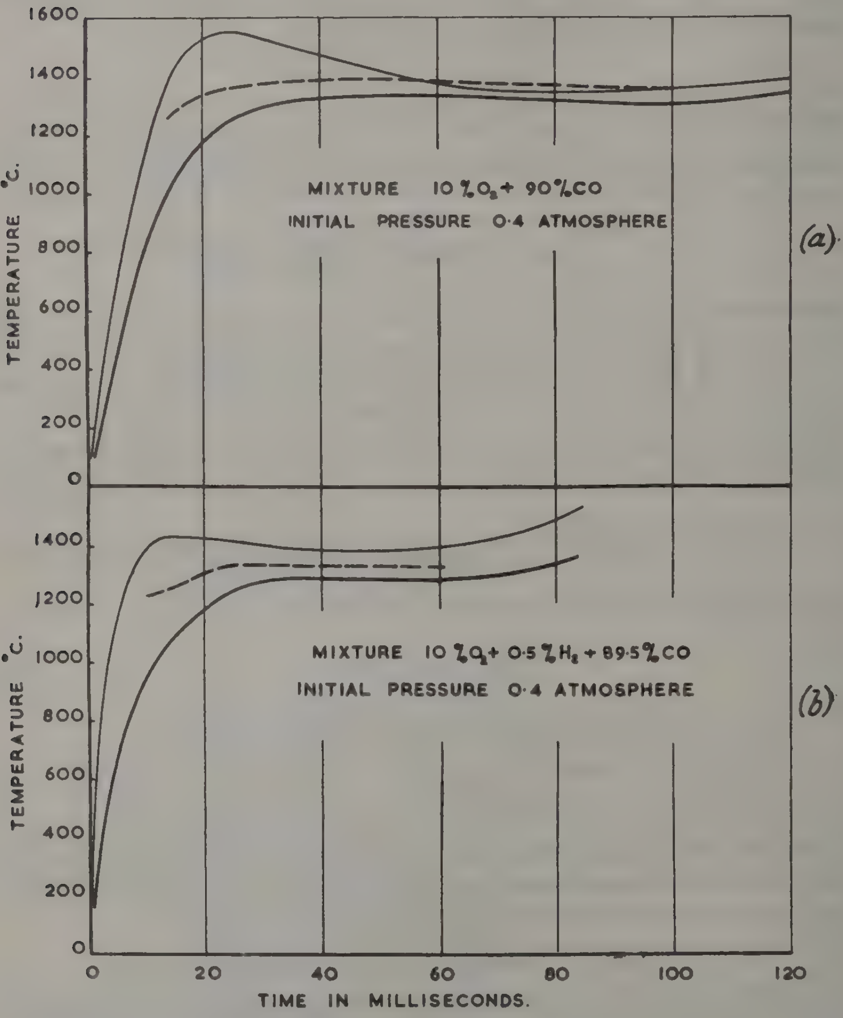
The left hand side of (13) is the rate at which O atoms are striking the wire per unit surface and may be determined from the catalytic heating curves, at any time t, hence n' can be found at this time t.

Explosions of a 10 *per cent* O_2+90 *per cent* CO *Dry Mixture.*

These explosions were carried out in the same spherical explosion vessel of 18 in. internal diameter and utilizing the same technique of determining rates of catalytic heating on plain platinum–rhodium wires as was described in a previous paper (Leah, Rounthwaite and Bradley 1950). In fig. 2 (*a*) are shown the plain and quartz-coated wire temperatures attained in an explosion of a dry, hydrogen free mixture of 10 per cent

O_2+90 per cent CO at an initial pressure of 0·4 atmosphere and an initial temperature of 15° C. These curves are reproduced from the previous paper as also is the full line curve of fig. 3 showing the variation of the rate of catalytic heating deduced from the temperature records of fig. 2 (*a*). The flame gas temperature obtained by correcting the quartz-coated wire

Fig. 2.

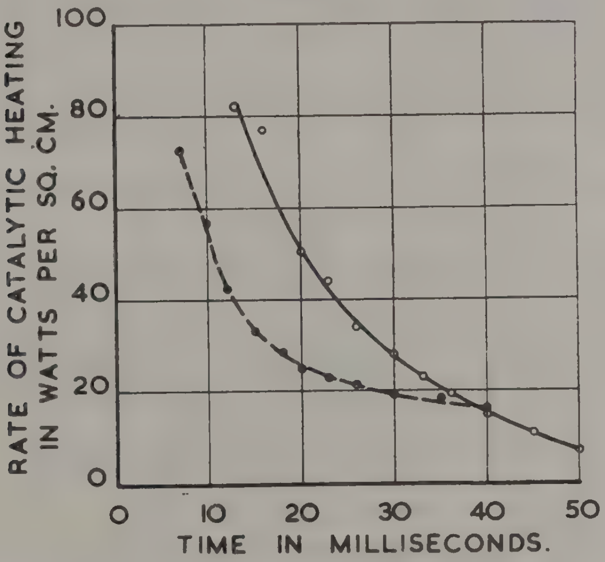


temperaure for radiation and heat capacity effects is shown in fig. 2 (*a*) by the broken line. It will be seen that the flame gas temperature is well below the ideal constant pressure temperature for this mixture, which is 1680° C., and that it remains reasonably constant, sufficiently so to be taken as constant in the subsequent diffusion analysis.

It is rather important to note that the temperature records are only of value in the present work during the early stages of the explosions whilst the gases are being burnt at very nearly constant pressure. After the so-called "pre-pressure" period the wire temperatures increase in consequence of the increase of flame gas temperature due to adiabatic compression resulting from the rising pressure. This temperature rise will be seen in the records reproduced in fig. 2 (*a*) and the portions of the curves beyond the point at which the rise commences have therefore been discarded in the subsequent analysis.

Taking the reaction $O+CO=CO_2+127$ kcals. to be proceeding exclusively on the plain wire it is possible to deduce the number of O atoms and CO molecules which must be reacting on the wire in order to release the amount of heat given by the curves of fig. 3.

Fig. 3.



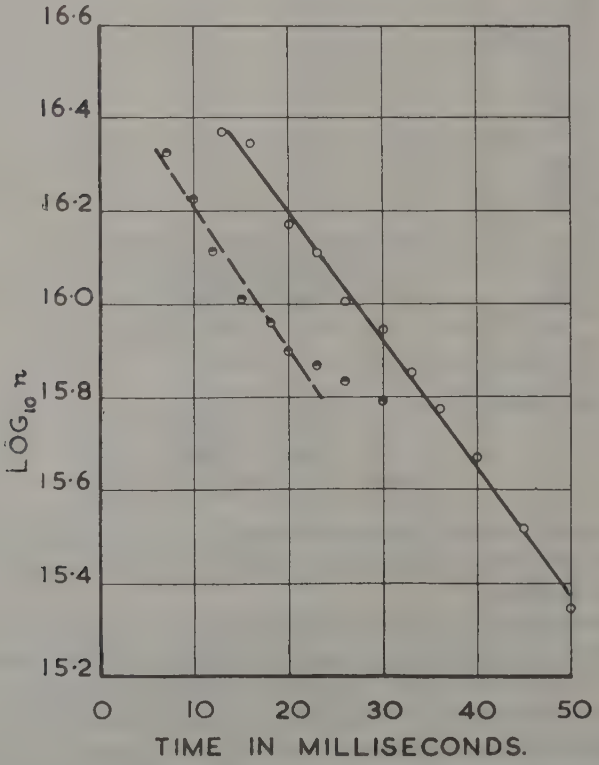
In accordance with the previous assumption that every atom of oxygen striking the wire reacts with CO the number of atoms reacting per sq. cm. per sec. is $D(\partial n/\partial r)_{r=a}$.

Thus, from equation (13), the concentration of O atoms, n', in the gases beyond the influence of the wire may be obtained. These values are given in Table I. for various times from the initial rise of temperature of the platinum wire. It should be noted that n' is the concentration of O atoms above the equilibrium concentration but as the latter is minute in the present explosions of mixtures containing excess CO, the values given may be taken to be the absolute concentrations. In using equation (13) the diffusion coefficient for the O atoms has been evaluated from the expressions of Chapman and Cowling (1939) using molecular data given by Rice (1940) and in the *Smithsonian Physical Tables* (1934).

TABLE I.

Time in milliseconds	O-atom concentration atoms per c.c. $\times 10^{-16}$	Percentage O-atoms by volume
13	2·34	1·30
16	2·22	1·25
20	1·47	0·80
23	1·30	0·70
26	1·01	0·55
30	0·88	0·50
33	0·70	0·40
36	0·59	0·35
40	0·46	0·25
45	0·33	0·20
50	0·22	0·10

Fig. 4.



If the rate of decay of the O atom concentration in the gas phase is given reasonably closely by equation (6) then $\log_e n' = gt + a$ constant. Plotting $\log_{10} n'$ against t for the explosion the full line curve of fig. 4

is obtained. It will be observed that for the greater part of the curve the experimental points lie very closely on a straight line, the slope of the line being proportional to g. The actual value of g obtained is 61 sec.$^{-1}$ at a mean gas temperature of 1360° C. and at 0·4 atmosphere pressure. The fact that the relationship between $\log n'$ and t is linear seems to lend strong support to the assumption which was made that the rate of decay of the O atoms was given closely be equation (V).

Explosions of this same mixture at higher initial pressure, up to 1 atmosphere, also show similar effects, the relationship between $\log n'$ and t always being linear in the early stages. Unfortunately it was found impossible to ignite a dry 10 per cent O_2+90 per cent CO mixture at pressures very much below 0·4 atmosphere and even when the mixtures did ignite at slightly lower pressures the flame speed was so low that convective rise of the inflamed gases caused them to rise out of the field of the wires. For this reason it was felt that a little hydrogen should be added to the mixtures in order to facilitate ignition at the lower pressures and to effect more rapid combustion. A new series of experiments was therefore carried out with a mixture consisting of 10 per cent $O_2+0{\cdot}5$ per cent $H_2+89{\cdot}5$ per cent CO.

Explosions of a 10 *per cent* $O_2+0{\cdot}5$ *per cent* $H_2+89{\cdot}5$ *per cent* CO *Dry Mixture.*

It was, of course, appreciated that the addition of hydrogen to the mixture must have a profound influence upon the reaction mechanism of the explosions. This case has been considered qualitatively by Semenov (1935), Hinshelwood (1940), Lewis and von Elbe (1934, 1938) and others. Additional reactions of the following types will be introduced :

$$\mathrm{H+O_2\rightarrow OH+O},$$
$$\mathrm{O+H_2\rightarrow OH+H},$$
$$\mathrm{OH+H_2\rightarrow H_2O+H},$$
$$\mathrm{OH+CO\rightarrow CO_2+H},$$
$$\mathrm{H+O_2+M\rightarrow HO_2+M},$$
$$\mathrm{CO+HO_2\rightarrow CO_2+OH}.$$

They have the effect of very materially increasing the rapidity of chain branching and the velocity of flame propagation but whether they should greatly effect the concentration of O atoms in the later stages of reaction it seemed impossible to predict. At the outset, therefore, explosions were carried out at an initial pressure of 0·4 atmosphere so that a direct comparison could be made with the hydrogen-free mixture previously used. The temperature records are shown in fig. 2 (*b*) together with the flame gas temperatures deduced from the quartz-coated wire temperatures. The corresponding catalytic heating curve is shown by the broken line of fig. 3. Comparison of fig. 2 (*b*) with fig. 2 (*a*), the latter referring to the hydrogen free mixture, and also comparison of the corresponding two curves of fig. 3, reveals that the rates of catalytic heating due to reactions on the plain wires are of the same order and decay in a similar manner in the two cases.

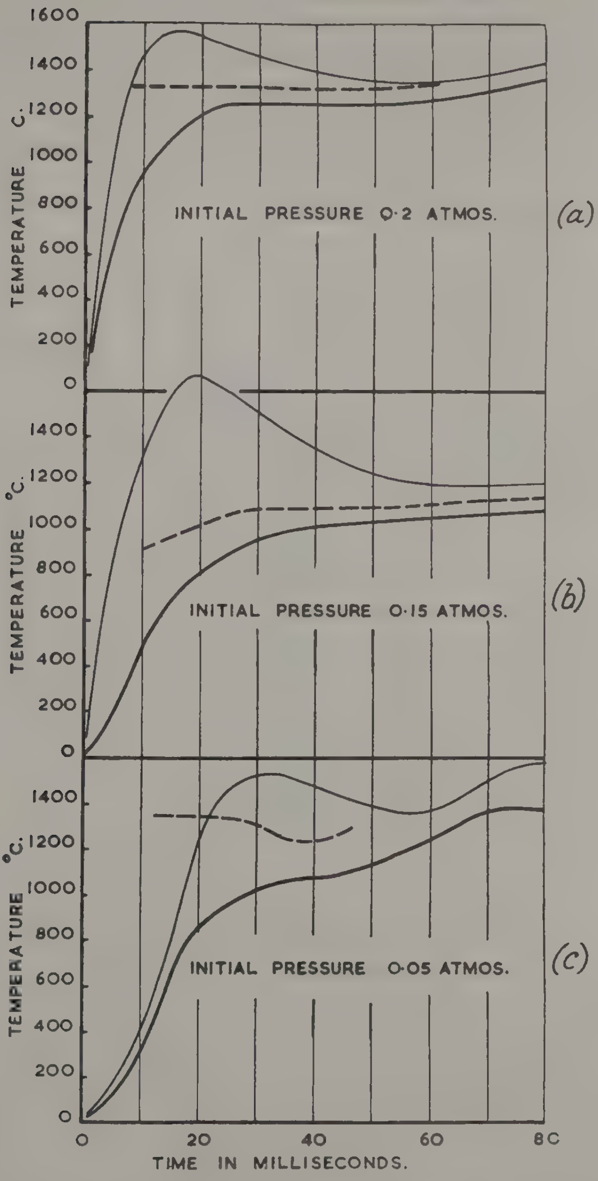
It is unlikely that the main catalytic reactions producing this heating of the plain wires can be due to any other reactions but those responsible for the heating with hydrogen free mixtures, for the hydrogen added is only present initially in small concentration. The possibility of H atom reactions on the wire surface cannot of course be excluded but it is felt that they play only a minor part in the catalytic heating of the wires. If these views are valid the analysis previously applied can once again be used to determine the atomic oxygen concentration in the gas phase. The results are shown on the $\log_{10} n'-t$ curve of fig. 4 by the broken line. The values of g agree almost identically and from this it might be inferred that the hydrogen addition has had little effect on the rate of decay of the atomic oxygen.

The addition of the small amount of hydrogen has been found to have no appreciable effect on the atomic oxygen concentration at any explosion pressure from 0·4 up to 1 atmosphere. These mixtures, unlike the hydrogen free mixtures, can be ignited at initial pressure down to as low as 0·05 atmosphere without any disturbance of the wire temperature records due to convective rise of the flame gases. At the lower pressures, however, an electric spark of greater intensity than was necessary at 0·4 atmosphere initial pressure had to be used in order to ignite the mixtures.

The wire temperature records for explosions at 0·2, 0·15 and 0·05 atmosphere initial pressure are shown in figs. 5 (*a*), (*b*), (*c*) respectively, the flame gas temperature being shown in each case by the broken line. Inspection of these temperature records, together with those given in fig. 2 (*b*) for the 0·4 atmosphere explosion, shows that the difference between the plain and quartz-coated wire temperatures increases with reduction of initial pressure, at least down to 0·15 atmosphere. The curves of fig. 6 which give the actual rates of catalytic heating on the plain wires at the different initial pressures, show that the rate of catalytic heating is somewhat greater as the pressure is reduced.

At the lowest pressure of 0·05 atmosphere the heating curve is of rather different form to those at the higher pressures, the rate at first increasing and then falling away again as in the other cases. The explanation of this phenomenn is almost certainly that with reduction of pressure the rate of chain branching is reduced, due to the lower frequency of collisions, and this has the effect of making the maximum concentration of atomic oxygen occur further behind the flame front (*i. e.* a greater time after the passage of the flame front over the wire). This heating curve has an obvious similarity to the curve of atomic oxygen concentration in the reaction zone for ozone explosions derived theoretically by Lewis and von Elbe (1934, 1938). In fact it is quite certain that had our technique been

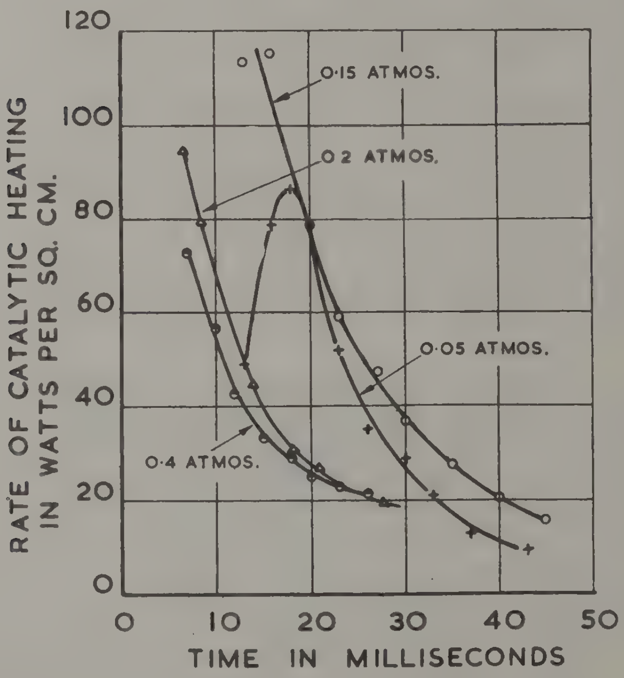
Fig. 5.



capable of recording the catalytic heating rates immediately after the passage of the flame front, all the curves shown in fig. 6 would have been found to pass through maxima and fall again as is observed at 0·05 atmosphere pressure.

From the diffusion equation the concentration of atomic oxygen at various times from the beginning of the rise in wire temperature has been calculated for each explosion pressure. The results obtained are given in Table II. being expressed in atoms per c.c. and also as percentages by volume of the total products of combustion. At each pressure the maximum concentration will be seen to be of the order of 2 per cent, a value which at these temperatures and pressures is far above the equilibrium value. For the explosion at 0·05 atmosphere initial pressure the values obtained may be considerably in error. This is because the chain branching reactions are so retarded at the low pressure that the maximum rate of catalytic heating on the wires does not occur in a very short time after the pasage of the flame front over the wire, as in the case of the higher pressure explosions. Thus the assumption made in deriving the expressions for calculating the concentration of O atoms, namely, that at time $t=0$, the concentration was a maximum in the gas phase and equal to n_0, will possibly introduce considerable errors. In addition, at 0·05 atmosphere initial pressure, the assumed rate of decay of the O atoms, $d[O]/dt=g[O]$, cannot be applicable during the earlier period as the rate of catalytic heating (and possibly the O atom concentration) is increasing.

Fig. 6.



In Table II., therefore, the results of the calculations for 0·05 atmosphere conditions are given only for the period after the peak of the catalytic heating curve and are not to be interpreted as anything more than very rough approximations. Of course, this does not in any way invalidate the conclusions drawn from the curve of catalytic heating which is quite independent of any assumed reactions occurring on the wires. Inspection of the catalytic heating curves of fig. 6 will show, however, that no serious errors due to this assumption will result in the case of explosions at higher initial pressures.

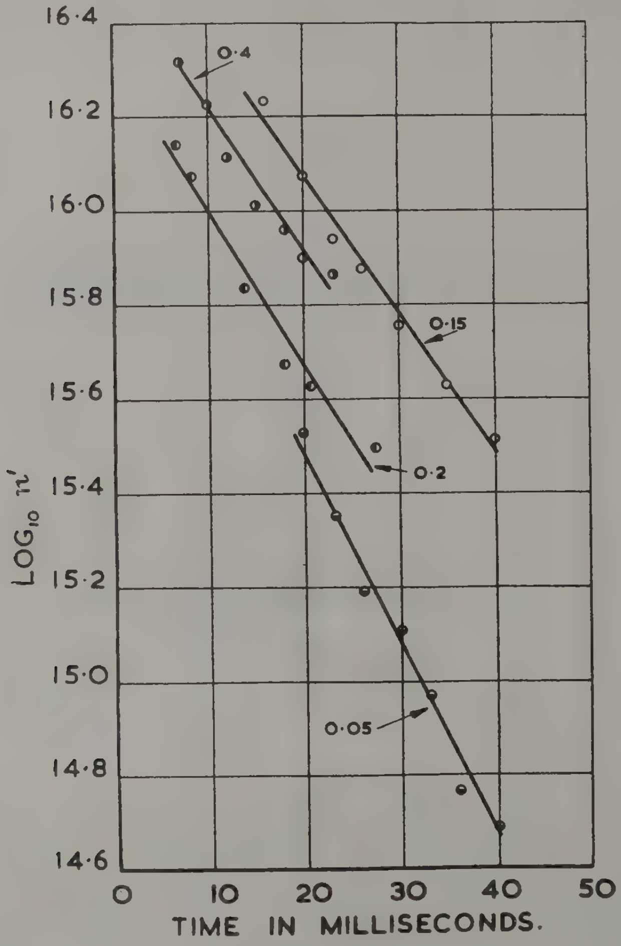
TABLE II.

Time in milli-seconds	Initial pressure in atmospheres							
	0·4		0·2		0·15		0·05	
	O-atoms per c.c. $\times 10^{-16}$	Per cent O-atoms by vol.	O-atoms per c.c. $\times 10^{-16}$	Per cent O-atoms by vol.	O-atoms per c.c. $\times 10^{-16}$	Per cent O-atoms by vol.	O-atoms per c.c. $\times 10^{-15}$	Per cent O-atoms by vol.
7	2·10	1·10	1·33	1·45				
10	1·69	0·90	1·00	1·10				
12	1·28	0·65	0·81	0·90				
15	1·03	0·55	0·61	0·65				
16	0·98	0·50	0·56	0·60	1·72	2·10		
18	0·91	0·50	0·48	0·50	1·45	1·80	3·69	1·60
20	0·79	0·40	0·43	0·45	1·19	1·45	3·42	1·45
23	0·73	0·40	0·36	0·40	0·89	1·10	2·25	0·95
26	0·69	0·35	0·34	0·35	0·75	0·90	1·54	0·65
30	0·62	0·35	0·29	0·30	0·57	0·70	1·28	0·55
33							0·93	0·40
35			0·25	0·25	0·43	0·50	0·66	0·30
40			0·23	0·25	0·33	0·40	0·49	0·20
43							0·42	0·20

In fig. 7 are shown curves of $\log_{10} n'$ against the time t for the explosions at initial pressures of 0·4, 0·2, 0·15 and 0·05 atmosphere. The calculated points appear to conform very well to the linear relationship which has been assumed although in all cases there is some divergence at the later periods. This is probably due to two factors, namely, the rising pressure and temperature resulting from adiabatic compression as the volume of inflamed gas increases and also the slowing down of the reaction as the O atom concentration approaches the equilibrium value, or the pseudo-equilibrium value if the presence of excited molecules of CO_2 disturbs the normal equilibrium. In addition it must be noted that at the later periods, the difference between the plain and quartz-coated wire temperatures becomes quite small and any errors of measurement will have a correspondingly larger effect.

The values of g at each of the different pressures, as derived from the gradient of the $\log_{10} n'-t$ curves, are given in Table III. It will be seen from this Table that the flame gas temperatures are not very different at the various pressures and also that the values of g are of much the same value at each pressure. This is somewhat strange for it would appear to indicate that the decay of atomic oxygen was due to a first order reaction. Now if reaction (3), namely, $O+CO+M\rightarrow CO_2+M$ was solely responsible for the disappearance of the O atoms the reaction rate would be

Fig. 7.



$$\frac{d[O]}{dt}=-k[O][CO][M],$$

where k is the velocity constant for the forward reaction, or

$$g = -k[\mathrm{CO}][\mathrm{M}].$$

Hence g should be proportional to the square of the pressure at any one temperature. The temperature variations shown in Table III. are by no means sufficient to account for the fact that g appears to be independent of pressure. It must be concluded that the reactions occurring in the gas phase are not, in the overall process, due solely to a third order reaction. They must in fact be of such a character as to give the appearance of a much lower order reaction. So little appears to be known of the kinetics of the explosive reaction of CO and O_2 that a possible reaction scheme cannot readily be predicted to harmonize with these observations and we prefer, at this stage, to leave the question open.

Attention has previously been directed to the fact that the flame gas temperature derived from the quartz-coated wire temperatures are considerably less than the ideal values. David and co-workers (1944) have also observed this defect of temperature (in their experiments radiation loss corrections were not applied and they assumed that the quartz-coated wire temperatures gave reasonably accurate estimates of flame gas temperature) and they ascribed it to abnormal electronic excitation of a proportion of the CO_2 molecules, these being assumed to possess excess intra-molecular energy which they called " latent energy ". There does in fact seem to be little doubt that excited CO_2 molecules, possibly in a metastable electronic state, are present in the flame gases resulting from explosions of CO–O_2 mixtures. The only doubt is whether they are present in sufficient concentration to account for the whole of the latent energy, a doubt which Egerton and Ubbelohde (1934) have previously expressed.

TABLE III.

Initial pressure atmospheres	Mean gas gas ° C.	g sec^{-1}
0·4	1280	69
0·2	1320	74
0·15	1080	67
0·05	1320	95

The atomic oxygen which, from the present work, appears to remain in the flame gases will naturally account for a certain proportion of the total missing energy. A single example will suffice to show that this is only a minor effect. With the explosion at 0·4 atmosphere initial pressure, in which the flame gas temperature is highest and the total missing energy least for the explosions here considered, the atomic oxygen accounts for some 9·5 per cent of the heat of combustion of the mixture at its highest recorded concentration at a time of 0·007 second after the temperature rise commences and this leaves an unaccounted energy of 20 per cent. After the atomic oxygen has well nigh disappeared, at a time of 0·04 second, the unaccounted-energy is still some 20·5 per cent of the heat of combustion of the mixture. The small increase may be, and probably is, due to the radiation loss from the flame gases during the time interval of 0·033 second. Radiation, however, is unlikely to explain the whole discrepancy and the conclusion would seem to support David's views that the energy is held in the flame gases in " latent " form.

Gaydon (1942) has given an alternative explanation of the latent energy in terms of vibrationally excited molecules. According to Gaydon, the presence of water-vapour—and presumably hydrogen—in the CO flame gases should have a marked effect in reducing the magnitude of the vibrational excitation energy. From the present work, however, the explosions of mixtures containing hydrogen show a large amount of missing energy after the elapse of an appreciable time. This appears to discount the views of Gaydon as giving a complete explanation of the phenomenon of latent energy. It seems to us most probable that the true explanation is to be found in a combination of factors such as radiation, electronic excitation, vibrational excitation and the presence of atomic oxygen and like active species. At this stage the relative importance of the first three of these factors cannot be assessed but the final factor does not appear, from the experiments which have been described, to have large influence.

REFERENCES.

Carslaw and Jaeger, 1947, *Conduction of Heat in Solids* (Oxford : University Press).

Chapman and Cowling, 1939, *The Mathematical Theory of Non-Uniform Gases* (Cambridge).

David, *et al.*, 1944, *Proc. Inst. Mech. Eng.*, **151,** 236.

Egerton and Ubbelohde, 1934, *Nature*, **134,** 848.

Gaydon, 1942, *Spectroscopy and Combustion Theory* (Chapman and Hall).

Gaydon, 1946, *Trans. Faraday Soc.*, **42,** 292.

Hinshelwood, 1940, *The Kinetics of Chemical Change* (Oxford : Clarendon Press).

Jost and Müffling, 1938, *Z. Physik. Chem.*, Sec A, **181,** 208.

Kondratieva and Kondratiev, 1946, *Acta Physicochim, U.R.S.S.*, **21,** 1.

Kondratjew, 1937, *Acta Physicochim., U.R.S.S.*, **6,** 748.

Laidler, 1949, *J. Chem. Phys.*, **17,** 221.

Langmuir, 1921, *Trans. Faraday Soc.*, **17,** 621.

Leah, Smeeton, A., 1943, *Phil Mag.*, **34,** 795 ; 1947, *Ibid,*. **37,** 657.

Leah, Smeeton, A., Rounthwaite, C., and Bradley, D., 1950, *Phil. Mag.*, **41,** 468.

Lewis and von Elbe, 1934, *J. Chem. Phys.*, **2,** 537 ; 1938, *Combustion Flames and Explosions of Gases* (Cambridge).

Rice, 1940, *Electronic Structure and Chemical Binding* (McGraw-Hill).

Semenov, 1935, *Chemical Kinetics and Chain Reactions* (Oxford : University Press).

Smith, W. V., 1943, *J. Chem. Phys.*, **11,** 110.

Tanford, 1947, *J. Chem. Phys.*, **5,** 433.

Taylor and Lavin, 1930, *J.A.C.S.*, **52,** 1910.

XLIII. *The Cut-Off in the Energy Spectrum of Primary Cosmic Radiation.*

By P. J. D. Gething,
Department of Mathematics, Imperial College of Science, London *

[Received January 30, 1950.]

Summary.

Janossy's explanation of the cut-off in the observed cosmic ray energy spectrum, viz., that it is due to the action of the Sun's magnetic field, was originally based on Hale's estimate for the dipole moment of the Sun. A recent investigation by Thiessen, however, has led to a much smaller estimate. It appears that, if Janossy's calculations were corrected to allow for this new estimate, the calculated level of the cut-off would be greatly reduced, and consequently there should be a corresponding increase in the total intensity of incoming radiation observable at high magnetic latitudes. The possibility of an alternative explanation of the cut-off is considered, with particular reference to the theories of Milne and Fermi.

1. The energy spectrum of arriving primary cosmic radiation (assumed throughout this paper to consist entirely of protons) is approximately given by:

$$n(\mathrm{E})\,d\mathrm{E}=\mathrm{const}\,.\,d\mathrm{E}/\mathrm{E}^{\gamma}, \quad \ldots\ldots\ldots \quad (1)$$

for E greater than 2×10^9 eV. In this formula $n(\mathrm{E})\,d\mathrm{E}$ is the number of particles with energies in the range E to $\mathrm{E}+d\mathrm{E}$. The exponent γ is estimated to be about 2·8. The fact that no radiation is reaching us of energy less than 2×10^9 eV., at least down to 3×10^7 eV., has been established experimentally by observations near the North Geomagnetic Pole (Carmichael and Dymond 1939).

The most generally accepted explanation of this cut-off has been the suggestion due to Janossy (1937), that particles arriving from outside the solar system are deflected by a permanent dipole magnetic field due to the Sun. Particles with momenta less than a certain critical value P_1 would then be unable to cross the Earth's orbit, whereas those with momenta between P_1 and a second critical value P_2 would only be able to arrive at the Earth from certain directions. Particles with momenta greater than P_2, however, could arrive from any direction.

---

* Communicated by E. A. Milne.

The general theory is complicated, but in the equatorial plane of the Sun's dipole field, which is assumed for simplicity to coincide with the plane of the Earth's orbit, the above critical momenta are:

$$P_1=(3-2\sqrt{2})eM/d^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$P_2=eM/d^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where e is the charge on the proton, M is the Sun's magnetic moment and d is the distance of the Earth from the Sun.

The cut-off directly observed refers to the momentum spectrum, and occurs at 2×10^9 eV./c. Taking this value for P_1, it follows from (2) that the corresponding magnetic moment of the Sun should be $M=10^{34}$ e.m.u., implying a field intensity on the Sun's surface of about 50 gauss at the poles.

2. When Janossy originally made the above calculations this result appeared to be in good agreement with more direct measurements of the solar field, particularly the observations of Hale (1913). Recently, however, Thiessen (1949) has re-investigated the Sun's field and claims that at the poles it amounts only to about 1·5 gauss, with a probable error of 0·5 gauss.

On Janossy's theory, therefore, the field required to account for the observations of Carmichael and Dymond would be from 30 to 40 times greater than the field observed by Thiessen. The only way of reconciling both sets of observations with Janossy's theory is to suppose that in the last decade or so the Sun's field has decreased about thirty-fold. Such a change, of course, would also affect the total intensity of radiation observed on the Earth. This may be roughly estimated as follows. Let us assume that the cut-off in the momentum spectrum has in fact changed from 2×10^9 eV./c. to 5×10^7 eV./c. and that the energy spectrum (1), with $\gamma=2{\cdot}8$, may be extrapolated to cover the corresponding energy range. Owing to the action of the Earth's magnetic field, particles with momenta less than 2×10^9 eV./c. cannot reach low magnetic latitudes, so that there would be no effective change up to the old "threshold latitude" at about 55° N. (geomag.). Beyond this point, instead of remaining constant, the intensity would continue to increase up to a new threshold latitude which would appear to be at about 70° N. (geomag.). From here to the magnetic pole the intensity would remain constant, but at a level some 15 or 20 times higher than the previous value.

3. If these approximate calculations are correct, it seems unlikely that the effect of a change in the Sun's magnetic field of the order of magnitude implied by the results of Hale and Thiessen respectively could have escaped detection. However, the following sources of possible inaccuracy in the calculations should be noted :—

(i) Observations by Pomerantz and Vallarta (1949) suggest the possibility that the extrapolation below 2×10^9 eV. of the energy spectrum law (1), with γ unaltered, may not be justified. If, however,

γ were reduced, the threshold latitude would still change as predicted above, although the change in total intensity might be very much less marked.

(ii) Rosseland (1936) has criticized the common procedure of treating the Sun's field as that of a dipole. Theory,suggests that this hypothesis may lead to serious errors in some applications. The dipole field, however, is probably a good enough approximation to deal with the theory of the orbits of the cosmic ray particles, even if it is not a justifiable approximation near the surface of the Sun.

(iii) Formulæ (2) and (3) apply only in the equatorial plane. Moreover, we have considered only possible changes in the value of P_1, without reference to P_2. Detailed investigation of the cosmic ray orbits depends on numerical integration, but it is unlikely that the conclusions reached above would be drastically modified if more exact calculations were made.

(iv) Formulæ (2) and (3) depend on the assumption that the moment M does not vary with time. Elementary calculations show immediately that the deflection suffered by a particle due to the Sun's field in all but the last few days of its journey towards the Earth is completely negligible. Consider, for example, the field in the equatorial plane at a distance of 3×10^{15} cm. (*i. e.* about 200 astronomical units) from the Sun. If M were 10^{34} e.m.u., the intensity would be $3{\cdot}7 \times 10^{-13}$ gauss. Now a proton of energy 2×10^9 eV. moving in a uniform field of this strength would describe a circle of radius $2{\cdot}4 \times 10^{19}$ cm., or about 26 light-years. Since the field of the Sun is not uniform, but falls off as r^{-3}, and since the nearest stars (which may be magnetically active) are only a few light-years from the Sun, it is apparent that the deflection caused by the Sun beyond 3×10^{15} cm. may certainly be ignored. More detailed investigation would probably show that this limit could be reduced. A particle whose speed is very close to that of light takes about 28 hours to travel 3×10^{15} cm. in a straight line. Even allowing for the curvature of the orbits, no modification will be needed to the theory of these orbits unless the Sun's field is rapidly fluctuating, with significant variations in a matter of a few days.

4. Since Janossy's theory seems to present some difficulty if Thiessen's observations are correct, it is interesting to enquire whether any other explanation of the cut-off can be found. The only alternative possibilities which seem to have been considered so far are the following :

(i) Hamilton, Heitler and Peng (1943) suggested that primary radiation is, in fact, reaching the Earth below the cut-off value and is escaping detection. Janossy (1947) has criticized this idea, but his reasoning is difficult to follow. The only requirements of their theory is that particles with energies below the critical value should undergo one atomic collision process in the atmosphere, in which the energy is converted to some form in which it cannot be detected by the coincidence counters and ionization chambers; while above the critical value the collision process is presumably different, and the incoming energy is converted to a detectable form.

(ii) Using Milne's theory (1935) that cosmic radiation is primarily due to world-gravitation, George (1948) has obtained a theoretical energy spectrum of about the right form, *i. e.*

$$n(\mathrm{E})\,d\mathrm{E}=\text{const}\,.\,(1-m_0^2c^4/\mathrm{E}^2)\,d\mathrm{E}/\mathrm{E}^3. \qquad \ldots\ldots \quad (4)$$

This curve rises rapidly from $n(\mathrm{E})=0$, when $\mathrm{E}=m_0c^2$, (*i. e.* for particles of zero momentum) to a maximum at $\mathrm{E}=(\sqrt{5/3})m_0c^2$, which for protons corresponds to $\mathrm{E}=1{\cdot}2\times10^9$ eV. After this point the curve falls approximately as const./E^3. It therefore includes something closely analogous to a cut-off at its low energy end. Furthermore, it seems that there may also be a true cut-off at some point in the spectrum according to this theory. In Milne's model of the Universe, a particle may be ejected from a fundamental particle or nucleus (these nuclei are identified with the extra-galactic nebulæ) by some subsidiary mechanism, and then accelerated by the combined gravitational field of all the nuclei to a speed arbitrarily close to that of light in a finite time. If inter-galactic space is as transparent as Hubble (1942) claims, then on Milne's theory it would appear that most of the cosmic ray particles which reach us have their origin in other galaxies. Since the nearest extra-galactic system, namely the Large Magellanic Cloud, is about 86,000 light-years away, the vast majority of the cosmic ray particles arriving at the Earth would have been travelling for at least 100,000 years. If there is in fact a minimum distance that the particles must have travelled, the theory that they are accelerated by a gravitational mechanism would appear to imply a corresponding minimum energy which they must have acquired by the time they reach the Earth.

(iii) Fermi's theory (1949), that the cosmic ray particles acquire their high energies from collisions with the magnetic fields associated with drifting interstellar clouds, has recently attracted considerable attention. He points out that if a proton is to be accelerated by this mechanism, its initial energy must be at least 2×10^8 eV., otherwise the energy losses due to ionization will exceed the energy gains. Since the rate of gain of energy will be irregular, it may be that this figure must be multiplied by a factor of 5 or 10 to give the particle any reasonable chance of undergoing a net acceleration. If this is the case, then some other mechanism must be found to bring the particles up to such a high "injection energy," but Fermi's calculations on ionization losses may be capable of application to other theories of origin.

CONCLUSION.

5. Recent observations by Thiessen on the Sun's magnetic field suggest that Janossy's explanation of the cut-off in the energy spectrum of incoming cosmic radiation should be re-examined and compared with

possible alternative theories. Future observations of the Sun's field, in conjunction with observations on the cosmic ray intensity and threshold latitude, should provide a crucial test of Janossy's theory.

I should like to thank Dr. G. J. Whitrow for helpful discussions on this subject.

Note added in proof.—Dolbear and Elliot have recently reported (1950, *Nature*, **165,** 353) measurements made at high altitudes of the variations in cosmic ray intensity throughout the solar day. They find that the amplitude of the variations is considerably smaller than the value predicted on the basis of Janossy's theory. This provides further evidence against the acceptance of his theory, which is independent of Thiessen's observations considered above.

REFERENCES.

CARMICHAEL and DYMOND, 1939, *Proc. Roy. Soc. A*, **171,** 321.
FERMI, 1949, *Phys. Rev.*, **75,** 1169.
GEORGE, 1948, *Phil. Mag.*, **39,** 240.
HALE, 1913, *Astrophys. J.*, **38,** 27.
HAMILTON, HEITLER and PENG, 1943, *Phys. Rev.*, **64,** 78.
HUBBLE, 1942, *American Scientist*, **30,** 101.
JANOSSY, 1937, *Zeit. f. Physik*, **104,** 430.
JANOSSY, 1947, *Commun. Dublin Inst. for Adv. Studies*, Series A, No. 4, 16.
MILNE, 1935, *Relativity, Gravitation and World Structure, O.U.P.*, Chapter XII.
POMERANTZ and VALLARTA, 1949, *Phys. Rev.*, **76,** 1889.
ROSSELAND, 1936, *Theoretical Astrophysics, O.U.P.*, 219.
THIESSEN 1949, *Observatory*, **69,** 228.

XLIV. CORRESPONDENCE

On the Binary and Ternary Photofission of Thorium-232.

By E. W. Titterton and T. A. Brinkley.

[Received February 21, 1950.]

A recent experiment by Titterton and Goward (1949) has shown that uranium undergoes photofission into three charged fragments. It is therefore of interest to determine whether a similar process occurs in the photofission of thorium which, after uranium is the naturally occurring element most easily fissionable with γ-rays.

It was expected, from the work of Westöo (1948) that difficulty would be experienced in loading plates with thorium-232 atoms, since his experiments had shown these to be adsorbed at the surface of the emulsion. Preliminary experiments made with 100μ Ilford D_1 emulsions employing a variety of presoaking times and concentrations of $Th(NO_3)_4$ $4H_2O$ in water confirmed Westöo's result that the take-up of thorium-232 ions from the solution is slow compared with that for thorium X and thorium B.

However, preswelling of the emulsion layer by a long soaking in water followed by immersion in thorium nitrate solutions of much higher concentration than that used by Westöo, resulted in improvements in penetration of the ^{232}Th ions. If the concentration is increased too far the sensitivity and mechanical properties of the emulsion are impaired, and it is necessary to make a compromise on this account.

In the experiments to be described, the plates were presoaked in water for 20 minutes at 18° C., were loaded by immersion at the same temperature in a solution of 10 grms. of $Th(NO_3)_4 4H_2O$ in 100 ml. of water for 20 minutes, immersed for two minutes in absolute alcohol and dried in a warm air stream. Under these conditions the sensitivity of the emulsion was unimpaired, and the ^{232}Th α-particles (range 16μ) were observed throughout the emulsion layer, although they were not distributed uniformly in depth.

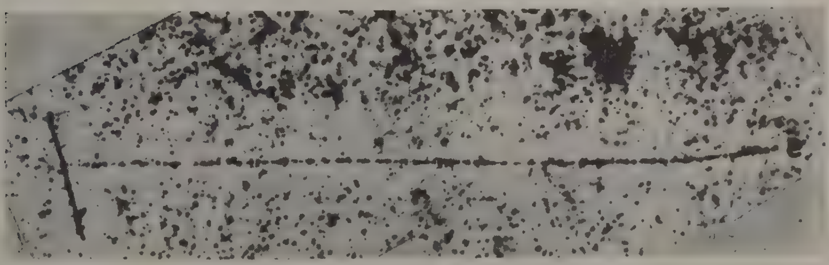
Plates loaded by this method were irradiated with a continuous spectrum of γ-rays of maximum energy 24 MeV. from the A.E.R.E. Synchrotron. Three irradiations, of 100, 150 and 180 R, were made and the resulting plates showed a fission density of $2\cdot5\times10^4$ per c.c. at the 150 R level, which, though small, was satisfactory for a preliminary investigation. Comparison of the exposed plates with identically treated, but unirradiated, control plates indicated some fading of tracks in the irradiated plates, especially deep in the emulsion. In an examination involving 2500 binary photofissions five cases of ternary fission involving the emission of a long range light fragment, probably an α-particle, have been observed. These events are listed in Table I. and a photomicrograph of one is given in fig. 1. In addition, there are seven cases where a third, short range, particle (which does not appear to be a recoil atom) occurs in the centre 5μ of the fission track. These events may be similar to those found in the slow neutron fission of ^{35}U where short range fragments of mass 12±1 have been reported (Dewan and Allen 1949).

TABLE I.

N	$f_1(\mu)$	$f_2(\mu)$	$F(\mu)$	$R_\alpha(\mu)$	E_α(MeV.)	ϕ(°)
1	15·1	10·2	25·3	160	18·0	93
2	14·7	10·3	25·0	133	16·3	77
3	11·6	11·6	23·2	83	12·6	71
4	12·9	11·0	23·9	91	13·1	71
5	13·1	12·4	25·5	68	11·2	82

N=number of event.
f_1=range of longer fission fragment.
f_2=range of shorter fission fragment.
$F=f_1+f_2$.
R_α=range of α-particle.
E_α=energy of α-particle.
ϕ=angle of emission of α-particle relative to long fission fragment.

Fig. 1.

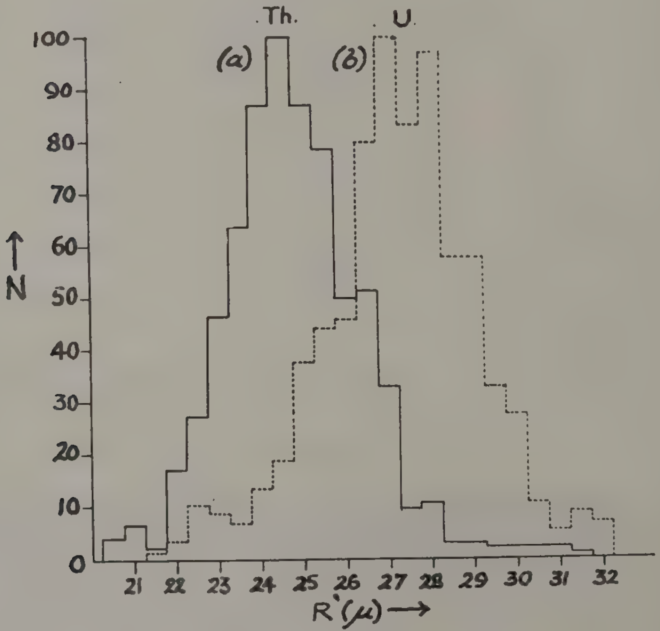


Photomicrograph of Event No. 2. Observer Miss P. A. Port. The two heavy fission fragments and the light, long range particle are wholly contained in the emulsion.

The possibility that the fissions observed are fast-neutron induced has been considered, but the fast neutron flux at the plates has been shown by experiment to be at least 25 times too small to account for the fission rate. Moreover, according to Tsien and Faraggi (1947) thorium-232 does not undergo tripartition with fast neutrons. Because of the long half-life of ^{232}Th, compared with that of its daughter products, the concentration of ^{232}Th atoms in the emulsion is far greater than that of any other member of the family. It is therefore reasonable to attribute the observed binary and ternary fission events to the action of γ-rays on ^{232}Th nuclei.

A number-range curve, fig. 2 (*a*), was determined for the photofission tracks and is compared with a similar curve, fig. 2 (*b*), for tracks formed by the slow neutron fission of ^{235}U in a D_1 emulsion under conditions of similar emulsion sensitivity. It appears that the energy release in the photofission of ^{232}Th is smaller than that in the slow neutron fission of ^{235}U. If it is assumed that the mass spectrum of the fission fragments in the two cases is similar, then from the velocity-range curves for the ^{235}U fission fragments given by Bohr (1941) it can be shown that, to a first approximation, the ratio of the kinetic energy released in the two cases is equal to the ratio of the squares of the corresponding ranges. Taking 24·5μ and 27·0μ as the mean ranges

Fig. 2.



Number-Range curves for binary fission tracks in D_1 emulsions :
(*a*) Photofission of Thorium,
(*b*) Slow-neutron fission of ^{235}U.

$$\frac{\text{Kinetic energy released in } ^{232}\text{Th photofission}}{\text{Kinetic energy released in } ^{235}\text{U slow neutron fission}} = \left(\frac{24\cdot5}{27\cdot0}\right)^2 = 0\cdot82$$

The measured value of the K.E. release in the slow neutron fission of ^{235}U is given by Brunton and Hanna (1949) as 151·5 MeV. and from this we derive 124±10 MeV. as the mean kinetic energy released in the photofission of ^{232}Th.

We wish to thank Mr. L. R. H. Metcalfe for help in setting up the experiment and for operating the synchrotron for these exposures.

Atomic Energy Research Establishment,
Harwell, Berks.
16th February, 1950.

E. W. Titterton.
T. A. Brinkley.

References.

Bohr, 1941, *Phys. Rev.*, **59,** 270.
Brunton and Hanna, 1949, *Phys. Rev.*, **75,** 990.
Dewan and Allen, 1949, *Phys. Rev.*, **76,** 181.
Titterton and Goward, 1949, *Phys. Rev.*, **76,** 142.
Tsien and Faraggi, 1947, *Compte-Rendus*, **225,** 294–6.
Westöo, 1948, *Ark. Mat. As. O. Fys.*, **34 B,** No. 22, 1–8.

A Temporary Increase in the Neutron Component of Cosmic Rays.

By N. Adams,
The Physical Laboratories, The University, Manchester.

[Received March 10, 1950.]

A recent report (Forbush, Stinchcomb and Schein 1950), has drawn attention to an abnormality in the cosmic ray intensity on November 19, 1949.

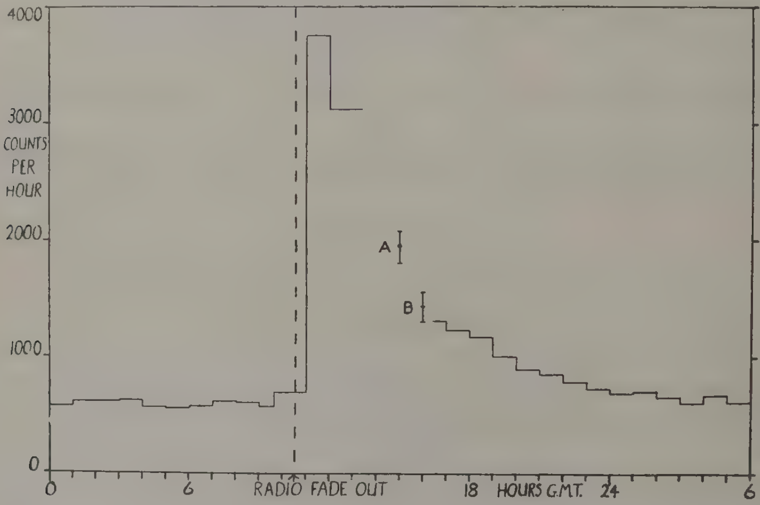
Increases of 43 per cent were found in the ionizing cosmic rays measured at sea level in geomagnetic latitudes 50° and 80°, and an increase of 180 per cent in the rays measured at Climax, Colorado, at an altitude of 11,000 feet. No increase was found at 11,000 feet on the equator. These increases were associated with a radio fade out at 10.33 G.M.T. At about the same time large increases were found in the cosmic ray intensities recorded by two independent systems in these laboratories.

One of these which records ionizing cosmic ray consists of three-fold Geiger–Muller counter telescopes pointing north and south at an angle of 45° to the horizontal (Elliot and Dolbear 1950). This recorded an increase in intensity reaching a maximum of about 12 per cent above normal between 11.00 and 12.00 hrs. Its counting rate had reverted to within 1 per cent of the normal by 14.00 hrs. The other is a neutron intensity recorder and the accompanying figure shows its response for the period 19th–20th November, 1949.

The neutron detector consists of boron–trifluoride filled proportional counters embedded in a " Pile " built of high purity graphite. Its detection efficiency is appreciable for incident neutrons of energies between thermal and perhaps 10 MeV., being greatest for energies 3 to 5 MeV.

The ordinates of the figure represent the observed average counting rate during successive hourly intervals. No hourly readings are available for the period between 13.25 and 16.27 hrs. when the equipment was being tested. However, counting rates taken for periods of a few minutes each and therefore of rather low precision, were obtained during this period at about 15.00 and 16.00 hrs. and are plotted in the figure with appropriate standard deviations as A and B respectively.

The barometric pressure for the period shown in the figure had extreme variations of ±0·1″ Hg about its mean value. Such variations would produce changes of about ∓2·5 per cent in the counting rate of the recorder in normal conditions.



From the moderate increase of counting rate (15 per cent) observed during the period preceding 11.00 hrs. and the enormous increase (550 per cent) in the period 11.00–12.00 hrs. (both compared with the counting rate earlier that morning) is seems reasonable to suppose that the onset of the disturbance occurred only a few minutes before 11.00 hrs.

After reaching a maximum value for the period between 11.00 and 12.00 hrs. the counting rate, so far as readings are available, decreased continually with time and was still significantly above normal in the first hours of November 20th. Thereafter the hourly counting rates showed their usual statistical fluctuation.

Except for the occasion described here no variation of counting rate of root mean square amplitude (averaged over periods of several weeks) greater than 0·3 per cent of the total rate has been found save for the effect of barometic pressure.

The author is indebted to D.W.N. Dolbear and H. Elliot for permission to publish the observations on the ionizing cosmic rays, described here, which have been made by them.

REFERENCES

FORBUSH, S. E., STINCHCOMB, T. B., and SCHEIN, M., 1950, *Bull .Amer. Phys. Soc.*, **25**, 15.

ELLIOT, H., and DOLBEAR, D. W. N., 1950, *Proc. Phys. Soc.* A, **63**, 137.

A Temporary Increase in the Neutron Component of Cosmic Rays.

By N. ADAMS and H. J. J. BRADDICK.
The University, Manchester.

[Received March 8, 1950.]

THE outstanding feature of the result reported above is the very large increase, at sea level, of neutron intensity as compared with ionizing cosmic radiation. This cannot be explained by a uniform increase in all types and energies of incoming cosmic rays.

Although fast neutrons from the sun could reach the earth with only a moderate loss due to decay, the increase cannot be due to the appearance of neutrons at the top of the atmosphere, for the increased rate persists after sunset. The rate was about 2½ times normal at Manchester sunset and was significantly above normal when the nearest point of the atmosphere directly irradiated by the sun, was hundreds of miles away.

The effect must therefore be ascribed to charged particles, and this conclusion is confirmed by the fact that no increase (in ionizing radiation) was observed on the geomagnetic equator. The particles concerned must have momenta between ca. $2 \cdot 10^9$ eV./*c*., the threshold value at Manchester and ca. $10 \cdot 19^9$ eV./*c*., the corresponding equatorial value. Particles at the lower end of this range would have rather low efficiencies as meson producers, and any mesons they produced would be completely stopped in the atmosphere, for the minimum ionization loss is ca. $2 \cdot 10^9$ eV. for a particle which traverses the atmosphere and there are no fluctuations in favour of individual particles. The absorption of neutrons in the atmosphere is of a different type—it takes place in a rather small number of nuclear encounters and is therefore exponential with depth. The altitude and barometric coefficients indicate that neutrons produced near the top of the atmosphere give rise to between 1/300 and 1/1000 of their number at sea level. For comparison, the mixed ionizing radiation at its maximum is about 50 times the corresponding sea level value. An increase in charged cosmic particles which favoured a comparatively narrow band above $3 \cdot 10^9$ eV. would therefore give rise to a large increase in sea level neutrons combined with a relatively small increase in sea level ionizing radiation.

XLV. *Notices of New Books and Periodicals received.*

A General Kinetic Theory of Liquids. By M. Born and H. S. Green. (Cambridge University Press). Price 10*s.* 6*d.*

In view of the concise and allusive style imposed on modern publications in the scientific journals, a book consisting only of a collection of papers with a brief appendix has necessary defects as an introduction to the subject of which it treats. The reviewer must then attempt to exhibit its relation to other work in the field, and to criticize its content in more detail than is usually required.

It has been usual to discuss the liquid state of matter in terms of models constructed in analogy with those found fruitful in the solid or gaseous states; and a good account of the equilibrium properties of normal liquids can be given by such models with reasonably few empirical parameters. This, however, can be done by any model which does not contradict the law of corresponding states, and it has proved much less easy to find any consistent picture of transport processes or to introduce quantum-mechanical effects.

A more fundamental approach is through the use of multiple distribution functions

$$f_h\,(x_1, \ldots x_h\,;\ \dot{x}_1, \ldots \dot{x}_h\,;\ t),$$

where f_h is the probability density at time t for finding different molecules at $x_1, x_2, \ldots x_h$ with velocities $\dot{x}_1, \ldots \dot{x}_h$ respectively. The book under consideration contains the contribution of Born and Green to this aspect of the theory, with its generalization in quantum mechanics. In the first paper, the classical distribution functions are introduced and continuity equations established. The introduction of a potential function establishes recurrence relations among the f_h.

For the case of thermodynamic equilibrium, these multiple distribution functions were introduced by Onsager (1933) and subsequently used by Kirkwood (1935, 1942). In this case the form of f_N, where N is the total number of molecules in the system, is known to be exp $[(F-H)/kT]$, where F is the free energy and H the Hamiltonian of the system. Provided the velocities only appear in the kinetic energy, it is possible to integrate over all values of all the velocities, obtaining multiple densities n_h. The h-molecule density, n_h, is proportional to

$$\exp\,[-(\text{potential of average force on } h \text{ molecules}/kT)],$$

where the average is taken over the canonical assembly. Kirkwood (1935) has suggested that if the potential energy is a sum of two-body terms, then in considering groups of more than two molecules we may assume that the average forces between pairs of molecules can be simply superposed. This leads in particular to an expression for n_3 in terms of n_2, which Born and Green assume valid in non-equilibrium states also and use to obtain an integral equation for the n_2.

The Kirkwood assumption is vital in the development of the theory. Born and Green give a rather misleading reference in the appendix to a result of Sabry quoted by Rodriguez (1949). The assumption is shown to be true for a solid in which the atomic equilibrium positions are on a perfect lattice and interactions are harmonic. It is obviously also true for a perfect gas, but an approximation valid in two extreme cases where the motion can be resolved in terms of normal coordinates simply related to the atomic coordinates is not necessarily so in less simple conditions.

In the second paper, Green insists on the kinetic character of the theory and its independence of statistical mechanics, so finds it necessary to use some form of the H-theorem. This meets the usual difficulty that the H-function, reasonably defined, is strictly constant in the motion of an unperturbed conservative dynamical system, no matter how improbable the initial state, so it is necessary to introduce thermal perturbation. This is represented by a simplified interaction with another similar system, which is no improvement on the use of the Gibbs assembly. A formal derivation of the equation of state from the integral equation for the n_2 follows.

For a corrected and more complete treatment of the equation of state the appendix refers to the paper of Rodriguez, which gives two computed isothermals for argon. These show a flattening at high pressures, which may represent the start of the liquid-solid transition region, and a cusp at the foot of the section which represents metastable states of the liquid under tension. This should be the point at which no activation energy is necessary to initiate a bubble: however, its significance depends on the credit one gives to the "unstable" section of the curve. Born and Green refer to forthcoming work of McLellan which should elucidate how far these results depend on the particular mathematical approximations employed.

The third paper in the book deals with transport processes. Born and Green relate the viscosity and thermal conductivity to the distorted distribution functions in the liquid in non-equilibrium conditions. The calculation of these distorted distributions is treated formally here by expansion in terms of the gradients of the intensive thermodynamic variable. Yang (1949 a, b) has discussed self and mutual diffusion in liquids, but none of this work has yet been reduced to numerical calculation.

In the non-equilibrium case, the physical significance of the distribution functions is not altogether clear. Onsager (1931) and Kirkwood (1946) have represented a system out of equilibrium by an assembly with given initial values, *e. g.* of the statistical space moments of intensive properties, or of the coordinates and momenta of specified molecules. This has led Kirkwood to a treatment of transport processes in which time-smoothed distribution functions of f_h obey an extended Langevin equation with friction constants calculable from molecular properties. Kirkwood, Buff and M. S. Green (1949) have obtained a reasonable numerical estimate of the shear viscosity of liquid argon by this approach. The relation between this approach and that of Born and Green is not quite obvious.

The second half of the book, also containing three papers, concerns the quantum mechanics of liquids. The physical system is described by v. Neumann's statistical operator, which is represented by the density matrix in the coordinate representation. Density matrices for smaller numbers of particles are formed from that for the whole system, as in the classical case.

In the first of these papers it is shown that from the equations of motion of the density matrices may be deduced equations formally identical with those of classical hydrodynamics. However, the trace of the stress tensor appearing in those equations is not proportional to the thermodynamic pressure derived from the free energy. This remarkable result has been challenged by de Boer (1949 a) and its basis further elucidated in a reply by Green (1949). It arises from the finite value of the trace of a set of commutators which appear in the expression for the thermodynamic but not the kinetic pressure. Similarly, the kinetic temperature appearing in the equation for heat conduction differs from the thermodynamic temperature.

The second paper contains a clear and instructive account of the relation between the statistical operator and thermal quantities. In the last paper, Green shows that quantum effects permit the rapid reversible interconversion of mechanical and thermal energy, and applies this to a qualitative explanation of anomalous transport phenomena in liquid helium II. In view of the observed difference in behaviour of He_3 and He_4 at low temperatures, it is unfortunate that the development contains no explicit reference to the statistics obeyed by the atoms of the liquid. It seems possible that since rotational states of low energy correspond to quasi-rigid rotations of comparatively large groups of atoms, the symmetry conditions modify the high-order density matrices corresponding to such groups (*cf.* Teller and Wheeler (1938)).

This book contains a valuable contribution to the kinetic theory of liquids. The reader who wishes to obtain a balanced view of the subject would do well to study it in conjunction with the published work of Kirkwood and the review article on dense gases by de Boer (1949 b). G. Wyllie.

References.

de Boer, J., 1949 a, *Physica*, **15,** 843.
de Boer, J., 1949 b, *Rep. Prog. Phys.*, **12,** 305.
Green, H. S., 1949, *Physica*, **15,** 882.
Kirkwood, J. C., 1935, *J. Chem. Phys.*, **3,** 300.
Kirkwood, J. C. and Boggs, E. M., 1942, *J. Chem. Phys.*, **10,** 394.
Kirkwood, J. C., 1946, *J. Chem. Phys.*, **14,** 180.
Kirkwood, J. C., Buff, H. and Green, M. S., 1949, *J. Chem. Phys.*, **17,** 338.
Onsager, L., 1931, *Phys. Rev.*, **38,** 405.
Onsager, L., 1933, *Chem. Rev.*, **13,** 73.
Rodriguez, A. E., 1949, *Proc. Roy. Soc.* A, **196,** 73.
Teller, E. and Wheeler, J. A., 1938, *Phys. Rev.*, **53,** 778.
Yang, L. M., 1949 a, *Proc. Roy. Soc.* A, **198,** 94.
Yang, L. M., 1949 b, *Proc. Roy. Soc.* A, **198,** 471.

The Mathematical Theory of Communication. By Claude E. Shannon and Warren Weaver. (Urbana, 1949: University of Illinois Press.) [Pp. 117.] Price $2.50.

The greater part of this volume is substantially a reprint of a paper on this subject by a well-known member of the Bell Telephone Laboratories, published in the Bell System Technical Journal in July and October, 1948. Only a condensed form of the remainder written by the second author has previously appeared; but portions of it might well be read first by those seeking a general view of the subject before embarking on a mathematical treatment.

The Atmospheres of the Earth and Planets, edited by Gerard P. Kuiper. (Published by University of Chicago Press, for whom Cambridge University Press act as agents.) [Pp. 366.] 17 plates. Price 60*s*. net.

This is the Report of a Symposium held at Chicago in September, 1947. Of the seventeen separate contributions, ten are concerned with various aspects of the atmosphere of the earth, and it mainly falls on Dunham and Kuiper to re-dress the balance with accounts of those of the planets. Nevertheless the intention, to present a comparative study of atmospheres, is genuinely realized, and results in a volume containing much factual material, especially on atmospheric spectroscopy and optics. F. C. F.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XLVI. *The Refractive Index and Classical Radiative Processes in an Ionized Gas.*

By K. C. Westfold,
The Queen's College, Oxford*.

[Received March 10, 1950.]

Abstract.

The rôle of the refractive index in the classical theory of radiation in an ionized gas is investigated. The Lorentz formula for the macroscopic absorption coefficient is recovered from microscopic considerations, and the heuristic modification of the formula for the emissivity obtained in a previous paper is justified. The consequent modification of the free-free transition probabilities is indicated.

§1. Introduction.

So long as astronomical observations were confined to frequencies greater than the infra-red, astrophysicists have been able to interpret them in terms of a microscopic theory of radiative processes in a medium whose refractive index is always unity.

In recent years observations have been extended to radio frequencies which, although they allow an adequate interpretation within the classical theory of radiation, require a detailed study of the part played by a refractive index that is a function of the frequency of the radiation.

It is generally accepted that the solar atmosphere and interstellar space are populated by fully ionized atoms. Such a medium profoundly affects the propagation of radio-frequency radiation.

The classical theory of radiation in an ionized gas in the absence of magnetic fields has been considered in a recent paper by Smerd and Westfold (1949)†. They found it necessary to introduce the refractive index as a factor in the denominator, in building up the macroscopic absorption coefficient from the microscopic transition probabilities obtained by ignoring the refractive index, and again as a factor in the numerator of the formula for the emissivity.

The first was necessary to secure agreement with the classical Lorentz formula and the second so that the resulting formula for the intensity divided by the refractive index should agree with the Planck formula.

In this paper we shall justify this procedure by investigating the microscopic processes of absorption and emission, while taking into account the electromagnetic scattering by the other particles of the

---

* Communicated by the Author.

† This paper will be subsequently referred to as S.W.

medium. It is this latter effect which constitutes the macroscopic refractive index. Although many of the results obtained are not essentially new, it is necessary to state them in a form that shows their dependence on the refractive index explicitly.

§2. ELECTROMAGNETIC SCATTERING AND THE REFRACTIVE INDEX.

The relation between electromagnetic scattering and the refractive index has been considered extensively by Darwin (1924) and Hartree (1928, 1931). In an ionized gas, scattering, absorption and emission associated with the motions of the ions are negligible compared with those associated with electron motions.

The mean motion of a free electron, of charge e and mass m, in an electromagnetic field ($\mathbf{E}$, $\mathbf{H}$) is determined by the equation (see *e.g.* Burkhardt (1950))

$$m(\dot{\mathbf{v}}+\nu\mathbf{v})=e\left(\mathbf{E}+\frac{1}{c}\mathbf{v}\times\mathbf{H}\right),$$

where $\mathbf{v}$ is the mean velocity induced by the field. The units are Gaussian. The term $m\nu\mathbf{v}$ represents the average damping force due to elastic collisions, principally with ions. On free-path considerations ν is taken to be the collision frequency, but the kinetic-theory derivation of the absorption coefficient in S.W. indicates that $\frac{3}{4}\nu$ should be the collision frequency. Provided $v\ll c$, the velocity of light *in vacuo*, the second term on the right side may be neglected, leaving the linear relation

$$\dot{\mathbf{v}}+\nu\mathbf{v}=\frac{e}{m}\mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

Now under the condition $v\ll c$ the scattered field due to the induced motion of the electron is equivalent to that of an elementary current of strength $e\mathbf{v}/c$. All the other electrons are similarly affected and their coherent scattered fields are together equivalent to that of a distribution of induced current of density

$$\mathbf{I}=ne\mathbf{v}/c, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

where n is the electron density. Moreover this is the only source distribution of the field. The scattered field must, therefore, be identical with the inducing field and satisfy Maxwell's equations,

$$\left.\begin{aligned}\operatorname{curl}\mathbf{E}+\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t}&=0,\\ \operatorname{curl}\mathbf{H}-\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}&=4\pi\mathbf{I}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

The relations (2.1)–(2.3) determine the field in terms of one of the vectors. Since they are linear, it is sufficient to confine our attention to monochromatic fields of frequency

$$f=\omega/2\pi, \qquad (2.4)$$

such that the field vectors are the real parts of the complex vectors

$$\mathbf{E}=\mathbf{E}' \exp(-i\omega t), \text{ etc.} \qquad (2.5)$$

Then we find

$$\left.\begin{aligned} &\mathbf{I}=\sigma_f\mathbf{E}, \\ \text{where} \qquad &\sigma_f=i\frac{ne^2}{m\omega c}\frac{1}{1+i\nu/\omega} \end{aligned}\right\} \qquad (2.6)$$

is the complex conductivity,

$$\mathbf{H}=-i\frac{c}{\omega}\operatorname{curl}\mathbf{E}, \qquad (2.7)$$

and **E** is subject to the condition

$$\left.\begin{aligned} &\left(\nabla^2+\frac{\omega^2}{c^2}q_f^{\,2}\right)\mathbf{E}=0, \\ \text{where} \qquad &q_f^{\,2}=1+4\pi i\frac{c}{\omega}\sigma_f. \end{aligned}\right\} \qquad (2.8)$$

The refractive index and absorption coefficient follow directly from consideration of plane-wave solutions of (2.8). A wave travelling in the direction of the unit vector **n** has the form

$$\mathbf{E}=\mathbf{E}'' \exp\{-i\omega(t-q_f\mathbf{n}.\mathbf{r}/c)\}. \qquad (2.9)$$

Thus q_f is the complex refractive index which may be written

$$q_f=\mu_f+i\chi_f, \qquad (2.10)$$

where μ_f is the real refractive index and the absorption index, χ_f, determines the absorption coefficient. Both field vectors are normal to the direction of propagation, the magnetic vector being

$$\mathbf{H}=q_f\mathbf{n}\times\mathbf{E}. \qquad (2.11)$$

Writing

$$\begin{aligned} q_f^{\,2}&=M_f+iN_f \\ &=(\mu_f+i\chi_f)^2, \end{aligned} \qquad (2.12)$$

we find

$$\left.\begin{aligned} \mu_f^{\,2}&=\tfrac{1}{2}M_f\{\sqrt{(1+N_f^{\,2}/M_f^{\,2})}+1\}, \\ \chi_f^{\,2}&=\tfrac{1}{2}M_f\{\sqrt{(1+N_f^{\,2}/M_f^{\,2})}-1\}, \end{aligned}\right\} \qquad (2.13)$$

where, from (2.6) and (2.8),

$$\left.\begin{aligned} \mathrm{M}_f&=1-\frac{\omega_0^{\,2}}{\omega^2+\nu^2}, \\ \mathrm{N}_f&=\frac{\nu\omega_0^{\,2}}{\omega(\omega^2+\nu^2)}, \\ \omega_0^{\,2}&=4\pi ne^2/m. \end{aligned}\right\} \quad \ldots\ldots \quad (2.14)$$

The energy absorption coefficient is then

$$\kappa_f=2\chi_f\omega/c. \quad \ldots\ldots\ldots \quad (2.15)$$

These are the formula of the Lorentz theory.

If, as is usually the case, $\nu \ll \omega$, (2.13)–(2.15) yield the approximate results quoted in S.W.,

$$\left.\begin{aligned} \mu_f^{\,2}&=1-f_0^{\,2}/f^2, \\ \kappa_f&=\nu f_0^{\,2}/cf^2\mu_f, \\ \text{where}\qquad f_0&=\omega_0/2\pi. \end{aligned}\right\} \quad \ldots\ldots \quad (2.16)$$

§3. THE CROSS-SECTION FOR ABSORPTION BY A FREE ELECTRON AND THE ABSORPTION COEFFICIENT.

In this section we recover the Lorentz formula for the absorption coefficient by investigating the absorption cross-section of a free electron belonging to an ionized gas in a radiation field, taking account of the scattering properties of the medium.

The quantities in electromagnetic theory that are important in the theory of radiation are the energy density of the radiation field and the Poynting vector which corresponds to the radiative flux vector. Since a radiation field may be regarded as the linear superposition of monochromatic plane waves of all frequencies and travelling in all directions, we may restrict our consideration to plane-wave fields of the type (2.9). Burkhardt (1950) has pointed out that the energy of the radiation field together with the magnetic energy of the current distribution constitutes the total energy of the electromagnetic field. The energy density of the current distribution may be expressed in terms of the current density and the vector potential, **A**, of which it is the source function, as $-\frac{1}{2}(\mathrm{Rl}\,\mathbf{I})\,.\,(\mathrm{Rl}\,\mathbf{A})$. Then the energy density, U_f, and the Poynting vector, $\mathbf{S}_f$, are given by

$$\left.\begin{aligned} \mathrm{U}_f&=\frac{1}{8\pi}\{(\mathrm{Rl}\,\mathbf{E})^2+(\mathrm{Rl}\,\mathbf{H})^2+4\pi(\mathrm{Rl}\,\mathbf{I})\,.\,(\mathrm{Rl}\,\mathbf{A})\}, \\ \mathbf{S}_f&=\frac{c}{4\pi}(\mathrm{Rl}\,\mathbf{E})\times(\mathrm{Rl}\,\mathbf{H}). \end{aligned}\right\} \quad \ldots \quad (3.1)$$

Again, since the duration of the phenomena in which we are interested is large compared with the period of the light or radio waves observed, we may restrict ourselves further to consider time averages of U_f and $\mathbf{S}_f$ over a period $1/f$. These are given by (*cf.* Stratton 1941)

$$\left.\begin{aligned}\overline{\mathrm{U}}_f&=\frac{1}{16\pi}\{|\mathrm{E}|^2+|\mathrm{H}|^2+4\pi\mathrm{Rl}(\mathbf{I}\,.\,\mathbf{A}^*)\},\\ \overline{\mathbf{S}}_f&=\frac{c}{8\pi}\mathrm{Rl}(\mathbf{E}\times\mathbf{H}^*),\end{aligned}\right\}\quad . \quad . \quad (3.2)$$

where a modulus sign denotes the complex modulus and an asterisk the complex conjugate.

Since the net charge density is zero, the electromagnetic field may be specified completely by $\mathbf{A}$, whence

$$\mathbf{E}=i\frac{\omega}{c}\mathbf{A} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.3)$$

and $\mathbf{H}$ is given by (2.11). Substituting the results (2.6), (2.11) and (3.3) into (3.2), we get

$$\left.\begin{aligned}\overline{\mathrm{U}}_f&=\frac{1}{8\pi}\mu_f^2|\mathrm{E}'|^2,\\ \overline{\mathbf{S}}_f&=\frac{c}{8\pi}\mu_f|\mathrm{E}'|^2\mathbf{n}.\end{aligned}\right\}\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.4)$$

Thus

$$\overline{\mathbf{S}}_f=\frac{c}{\mu_f}\overline{\mathrm{U}}_f\mathbf{n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.5)$$

showing that the phase velocity, c/μ_f, is the velocity of energy propagation.

This result is contrary to an assertion made in S.W. that the velocity of energy propagation is $c\mu_f$, a statement which is true only for the propagation of a group of waves such as a pulse.

The average rate of absorption of energy by an electron from the field is

$$\begin{aligned}\overline{\mathrm{W}}_f&=e(\mathrm{Rl}\,\mathbf{v})\,.\,(\mathrm{Rl}\,\mathbf{E})\\ &=\tfrac{1}{2}e\mathrm{Rl}(\mathbf{v}\,.\,\mathbf{E}^*),\end{aligned} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.6)$$

as in (3.2). From (2.2) and (2.6) we find

$$\overline{\mathrm{W}}_f=\frac{e^2}{8\pi m}\,\frac{\nu}{f^2+\nu^2/4\pi^2}|\mathrm{E}'|^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.7)$$

It is clear from (3.7) that there is an overall extraction of energy from the field only by virtue of the collisions the electron makes with the other particles of the medium. During part of any period the rate of absorption of electromagnetic energy is negative. This represents the emission stimulated by the field. If no collision takes place, it is balanced by an equal amount of energy absorbed from the field during the other half of the period. Thus $\overline{\mathrm{W}}_f$ is the net absorption over a period. It is, on the average, communicated to the ions with which an electron collides.

The cross-section for absorption by a free electron, a_f, is defined such that

$$\overline{\mathrm{W}}_f = a_f \overline{\mathrm{S}}_f. \qquad (3.8)$$

Then by (3.4) and (3.7)

$$a_f = \frac{e}{\pi c m \mu_f} \frac{\nu}{f^2 + \nu^2/4\pi^2}. \qquad (3.9)$$

The absorption coefficient is now simply

$$\begin{aligned} \kappa_f &= n a_f \\ &= \frac{\nu f_0^2}{c f^2 \mu_f} \frac{1}{1 + \nu^2/4\pi^2 f^2}. \end{aligned} \qquad (3.10)$$

This is identical with the exact result obtained from (2.13)–(2.15).

§4. The Emmissivity.

There remains to show that scattering introduces a factor μ_f in the formula for the emissivity, η_f.

In S.W., Kramers's procedure of making a Fourier analysis of the total radiation, Q, emitted during a free-free electron-ion encounter was followed, giving Q_f such that

$$\mathrm{Q} = \int_0^\infty \mathrm{Q}_f \, df. \qquad (4.1)$$

Then Q_f was summed over all types of encounter taking place in a volume element over a small time interval, to give the emissivity. The first part of this analysis will have to be done afresh, taking scattering into account from the outset.

The microscopic source of the field under consideration is an electron accelerating in a hyperbolic orbit about a positive ion. Its acceleration at any instant is $\mathbf{j}(t)$, and we may conveniently take the duration of the encounter as from $t = -\infty$ to $t = +\infty$. The Fourier transform of is

$$\left.\begin{aligned} &\mathbf{J}(\omega) = \int_{-\infty}^{\infty} \mathbf{j}(t) \exp(i\omega t)\, dt, \\ \text{whence} \qquad &\mathbf{j}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{J}(\omega) \exp(-i\omega t)\, d\omega. \end{aligned}\right\} \qquad (4.2)$$

The Fourier transforms of $\mathbf{E}$ and $\mathbf{H}$ due to $\mathbf{j}$ are likewise

$$\left.\begin{aligned} &\overline{\mathbf{E}}(\omega) = \int_{-\infty}^{\infty} \mathbf{E}(t) \exp(i\omega t)\, dt, \\ \text{whence} \qquad &\mathbf{E}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{\mathbf{E}}(\omega) \exp(-i\omega t)\, d\omega, \text{ etc.} \end{aligned}\right\} \qquad (4.3)$$

Then the Maxwell-Lorentz equations (2.1)–(2.3) transform to

$$\left.\begin{aligned} \operatorname{curl} \overline{\mathbf{E}} - i\frac{\omega}{c} \overline{\mathbf{H}} &= 0, \\ \operatorname{curl} \overline{\mathbf{H}} + i\frac{\omega}{c} q_f^2 \overline{\mathbf{E}} &= 0, \end{aligned}\right\} \qquad (4.4)$$

where q_f^2 is given by (2.8) and (2.6).

These equations are formally identical with those conditioning a monochromatic field of frequency $f=\omega/2\pi$ such that $\mathbf{E}'=\bar{\mathbf{E}}df$.

We may, therefore, write down the foımulæ for $\bar{\mathbf{E}}$ and $\bar{\mathbf{H}}$ is terms of $\mathbf{J}$ directly from the corresponding formulæ for the field vectors due to a harmonic Hertzian oscillator of moment

$$e\boldsymbol{\xi}=e\boldsymbol{\xi}'\exp(-i\omega t),$$

in a medium of dielectric constant q_f^2.

Only the acceleration field is significant and $\ddot{\boldsymbol{\xi}}$ corresponds to $\mathbf{J}$. Then at a position $\mathbf{R}$, with respect to the source, we have

$$\left.\begin{aligned}\bar{\mathbf{E}}&=\frac{e}{c^2\mathrm{R}}(\mathbf{J}\times\hat{\mathbf{R}})\times\hat{\mathbf{R}},\\ \bar{\mathbf{H}}&=q_f\frac{e}{c^2\mathrm{R}}\mathbf{J}\times\hat{\mathbf{R}},\end{aligned}\right\}\quad\ldots\ldots\quad(4.5)$$

where the circumflex denotes a unit vector.

The total contribution to the emissivity during an encounter is

$$\mathrm{Q}=\int_{-\infty}^{\infty}dt\int_0\mathbf{S}\,.\,\mathbf{n}\,d\mathrm{S},\quad\ldots\ldots\quad(4.6)$$

where $\mathbf{n}$ is the unit normal directed out from a closed surface surrounding the locality of the encounter. Now

$$\begin{aligned}\int_{-\infty}^{\infty}\mathbf{S}(t)\,dt&=\frac{c}{4\pi}\int_{-\infty}^{\infty}\mathbf{E}(t)\times\mathbf{H}(t)\,dt\\ &=\frac{c}{8\pi^2}\int_{-\infty}^{\infty}\bar{\mathbf{E}}(\omega)\times\bar{\mathbf{H}}(-\omega)\,d\omega,\quad\ldots\quad(4.7)\end{aligned}$$

by one of the Parseval formulæ for Fourier integrals. This is the analogue of the monochromatic relation (3.2).

It then follows from (4.1) and (4.6) that

$$\mathrm{Q}_f=\frac{c}{2\pi}\int_0\mathrm{Rl}\{\bar{\mathbf{E}}(\omega)\times\bar{\mathbf{H}}(-\omega)\}\,.\,\mathbf{n}\,d\mathrm{S}.\quad\ldots\ldots\quad(4.8)$$

By (4.5)
$$\bar{\mathbf{E}}(\omega)\times\bar{\mathbf{H}}(-\omega)=q_f\frac{e^2}{c^4\mathrm{R}^2}|\,\mathrm{J}\,|^2\sin^2\Theta,$$

where
$$\mathrm{J}\cos\Theta=\mathbf{J}\,.\,\hat{\mathbf{R}}.$$

Since the integral of $\sin^2\Theta$ over a complete solid angle is $8\pi/3$, we obtain finally

$$\mathrm{Q}_f=\frac{4e^2}{3c^3}\mu_f|\,\mathrm{J}\,|^2.\quad\ldots\ldots\ldots\quad(4.9)$$

This is the result S.W. (3.23) with the additional factor μ_f.

The determination of $|\mathrm{J}|^2$ in terms of the initial velocity, g, and the distance of the interacting ion from an asymptote of the orbit, b, is the same problem as in S.W. With the same assumptions, that $fb/g \ll 1$ and that there have been sufficient collisions to bring about a Maxwellian distribution of the velocities of the particles, we shall arrive at the heuristic formula S.W. (3.53) which it was our purpose to justify.

§5. THE TRANSITION PROBABILITIES.

Although the emission of radiation by an electron in a classical (g, b) orbit is not (apart from possible modifications to the orbit) affected by the presence of the surrounding electrons, we have seen that these have the effect of scattering the radiation so that the contribution, Q_f, to the emissivity, η_f, in the frequency range $(f, f+df)$ is reduced according to the factor μ_f.

To interpret this statement in terms of the quantum theory, we may say that the probability, α_{21}, of the spontaneous emission of a quantum hf in the range $(f, f+df)$ accompanying a switch from an orbit (g_2, b_2) to an adjacent orbit (g_1, b_1) is reduced by the presence of the surrounding electrons according to the factor μ_f. The identification is such that (S.W.(3.43) and (3.45))

$$\left.\begin{aligned} \mathrm{Q}_f df &= \alpha_{21} hf \\ \text{and} \qquad hf &= \tfrac{1}{2} m(g_2{}^2 - g_1{}^2). \end{aligned}\right\} \quad \cdot \; \cdot \; \cdot \; \cdot \; \cdot \quad (5.1)$$

The probability coefficient, β_{21}, for a stimulated emission in a transition 2→1, and the coefficient, β_{12}, for an absorption in the converse transition 1→2, are both inversely proportional to α_{21}. They are given in S.W. (3.40) and (3.41). Since they contain the factor $1/\mu_f$, the consequent kinetic-theory derivation of κ_f will yield the heuristic formula S.W. (3.15).

This quantum-theory interpretation differs from the details suggested in S.W., where the refractive index was associated not with the transition probabilities but with the number of transitions taking place per unit volume and time. The final formulæ for η_f and κ_f involving μ_f are of course unaffected.

ACKNOWLEDGMENT.

This work was carried out during the tenure of a studentship from the Commonwealth Scientific and Industrial Research Organization, Australia.

REFERENCES.

BURKHARDT, G., 1950, *Annalen d. Physik*, **5**, 373–80.
DARWIN, C. G., 1924, *Trans. Camb. Phil. Soc.*, **23**, 137–67.
HARTREE, D. R., 1928, *Proc. Camb. Phil. Soc.*, **25**, 97–120; 1931, *Ibid.*, **27**, 143–62.
STRATTON, J. A., 1941, *Electromagnetic Theory*, (New York: McGraw-Hill), Section 2.20.

XLVII. *Investigation of Soft Radiations.—VII. Energy Expenditure per Ion-pair for Slow Electrons in Various Gases.*

By S. C. Curran, A. L. Cockroft and G. M. Insch,
Department of Natural Philosophy, University of Glasgow *.

[Received March 20, 1950.]

Summary.

The proportional counter is applied to the problem of determining the energy expenditure per ion-pair formed in argon, methane, ethylene, air, nitrogen, oxygen and carbon dioxide. Two different methods are applied. In the first the nearly homogeneous X-rays of Copper 64 are employed and the measurement of the number of disintegrations per minute occurring in a known fraction of the source is compared with the ion current through the same tube due to the whole source. Tritium gas is used likewise, in known relative amounts, to give rise to ion-currents and counting rates of easily measurable magnitudes. Hence the energy expenditure is obtained for the photo-electrons produced by Kα X-rays and for the slow β-rays of tritium (average energy=5·7 Kev). The advantages obtained by using slow β-rays in counters are stressed.

Introduction.

The experimental determination of the energy expenditure per ion pair formed by different ionizing particles at various velocities has been the subject of many investigations (Compton and Allison 1936). Much of the earlier data was inaccurate owing to the limitations of the techniques available. There is however surprisingly little data directly obtained for electrons of known energy and which aims at reasonable standards of accuracy. Reviews of parts of the subject have been given by Gray (1944) and Binks (1936) and some more recent values are tabulated in MDDC-1017. We note particularly that investigations with mono-energetic electrons have been made by Eisl (1929) and Gerbes (1935) for air and by Nicodemus (1946) for argon. The accurate determination of the energy expenditure is a matter of considerable importance and the work described below was undertaken in an effort to obtain accurate values for electrons in a number of different gases. In particular it was essential that values should be measured for the gases employed in proportional counting tubes since the resolving power of the instrument was involved (Curran, Cockroft and Angus 1949). It was realized that the proportional counter offered a new and powerful tool for the accurate absolute measurement of energy expenditure for slow electrons.

---

* Communicated by the Authors.

EXPERIMENTAL METHODS.

1. *Homongeneous X-rays.*

Studies of X-radiation performed with proportional tubes (Curran, Angus and Cockroft 1949) showed that they could be used to measure accurately the average energy of the quanta and to count the number of such quanta with small error. In principle therefore if the flux of quanta entering the tube is n-sec. and the average energy expended by the quanta is E electron-volts (obtained from the photo-electron and Auger electron liberated in the gas), the rate of production of ion pairs is $n\mathrm{E}/\mathrm{V}$. per sec. where V is the energy expended in producing each ion pair. If the tube is used to measure n and E by operating it as a proportional counter, and if it is then operated as an ionization chamber by reducing the applied voltage below the value for onset of gas multiplication, the current i produced by the ions can be measured. Hence since

$$i=n\mathrm{E}/\mathrm{V}. \qquad (1)$$

the value of V can be calculated. In actual practice the determination of V is not quite so straightforward as suggested here. Thus counting rates n of the order of 50,000/min. could be easily measured but the electrometer valve circuit used to observe the absolute value of i was not sufficiently sensitive to give the corresponding current i accurately. The source of radiation can however be adjusted by a known fraction upwards or downwards in intensity. If a fraction f of the source gives n quanta/sec. then we have

$$fi=n\mathrm{E}/\mathrm{V}. \qquad (2)$$

To make the current measurement accurate and stable a value of of $f\sim 1$ per cent was usually arranged.

2. *Beta-rays of Tritium.*

The use of X-rays as the primary agents responsible for the ionizing electrons is not ideal in practice since the fractional absorption of these in the gases investigated makes it necessary to maintain the same pressure throughout the experiment. Moreover the measurement of n involves the proper quantitative operation of the counter with all of the gases employed. It is not possible to obtain good plateaus on the characteristic curves of tubes filled with such gases as air, oxygen or carbon dioxide. For this reason an alternative proceedure was adopted. This depended on the use of the slow β-rays of tritium gas, H^3, as the ionizing electrons. A small quantity of tritium was taken and the rate of disintegration measured. It is to be noted that any filling gas with good counting features (*e. g.* argon+methane mixture) could be used in this part of the experiment. Indeed a suitable Geiger tube might perform satisfactorily. Then a much larger quantity of tritium, together with the gas or mixture of gases being investigated, was used to fill the counter which was subsequently operated as an ionization chamber. The larger quantity of tritium was of course accurately related to the smaller quantity used when counting was done. The ratio of the two quantities was either determined by measuring volumes of the gas or by measuring pressures.

The average energy of the β-rays forming the spectrum of tritium is accurately known and recent work with the proportional tube (Curran, Angus and Cockroft 1949) shows that it is 5·7 keV. (See also Slack *et al.* 1949) and Jenks *et al.* 1949). Clearly this second method offers some considerable advantages. Any gas which gives a saturation current in an ionization chamber can be investigated. Thus air, oxygen and other gases with appreciable electron-attachment coefficients are not excluded. Again the absolute value of V can be measured for one gas and the values of V for other gases obtained by using definitely related quantities of tritium in each gas in turn. It is to be noted that the partial pressure of the tritium can be kept very small and below a value likely to affect the accuracy of the results. Moreover the short range of the β-rays means that very little error ($\sim$1 per cent) is involved in assuming that all of their energy is expended in the gas even when chambers of moderate dimensions ($\sim$2 in. diameter) are employed and pressures $\sim$1 atmosphere. The transparency problem involved with X-rays is thus eliminated.

ABSOLUTE VALUE OF V FOR ARGON.

1. *X-ray Method.*

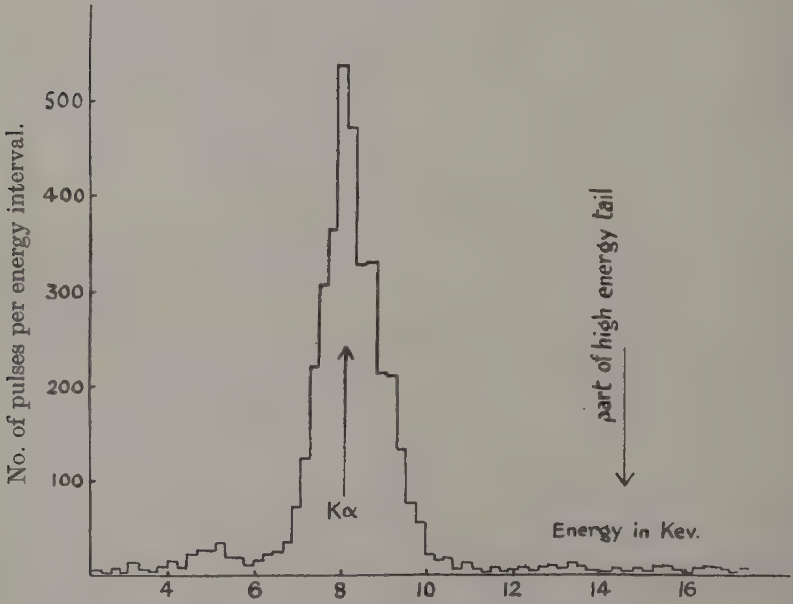
The choice of argon as the gas receiving most attention is justified in view of its widespread use in counters, ion chambers and cloud chambers. It was decided at the outset that the accurate measurement of the fraction f (equation (2) above) was of crucial importance. Various procedures could have been adopted but it seemed best to the authors to use radioactive decay. Fortunately a suitable source of soft X-rays exists in the form of Cu^{64} and the half-life (12.8 hrs.) was such that a period of only 3 days separated the measurements of ion-current and counting rate.

The small piece of radioactive copper was rigidly attached to the aperture of a proportional counter (4 in. diameter, 20 in. long) filled to atmospheric pressure with a 60/15 mixture of argon and methane. The ionization current through the tube was measured under conditions assuring good saturation. The source was allowed to decay and after a suitable interval the rate of counting was observed. The plateau obtained for the counter was fairly good although not entirely free from ambiguity. At the lower edge of the plateau the counting rate was higher than on its main stretch but it was thought that this was a spurious effect. After measurement of the counting rate the spectrum of the pulses was analysed as described elsewhere (Curran, Angus and Cockroft 1949). A good prominent peak corresponding to the Kα radiations of nickel was found (as shown in fig. 1) but the high-energy tail of low intensity produced by the annihilation γ-radiation from the source (which was covered with polythene to absorb β-particles) formed a large contribution to the total energy. The 60/15 mixture was not too fortunate a choice in this instance and a higher total pressure of argon would have been preferable.

The spectrum could be analysed over the entire energy range however, and it was found that $\overline{E}$, the average energy expended by the quanta in the gas was 12·7 keV. With this value, V for argon/methane in the ratio of 60/15 by pressure was 29·5 eV.

Regarding the experimental error this is difficult to assess. No account is taken of end-effect but absorption of the X-rays in the argon renders this less disturbing than usual—the source was placed at the centre of the tube. It is more likely that the total count was achieved than that complete saturation of the current was reached. (There was small but definite slope on the curve for current as a function of voltage.) In this way the authors consider that it is probably safe to express V in the form.

$$V=29{\cdot}5^{+0{\cdot}5}_{-1{\cdot}0}\,\text{eV}.$$



X–ray and γ–ray spectrum of Copper 64.

2. *Use of Tritium.*

Tritium was mixed with a little hydrogen as diluent and carrier. Experience gained in various attempts to define the fraction *f* which was counted led the authors to the conclusion that it was best to have two vessels identically constructed, one which acted as a counter and one which performed as an ion chamber. A supply of radioactive gas was prepared and a known fraction (by volumes) drawn off. This could be diluted further if required. The gas mixture containing the smaller activity was passed into the counter and the larger quantity into the ion chamber. The counter was then filled with a suitable counting mixture (frequently 7 cm. of CH_4 70 cm. of argon) while the ion chamber was filled with the main quantity. Usually the two vessels were very long (72 in.) and of relatively small diameter ($1\frac{1}{4}$ in.). The counter contained a thin wire (0·1 mm. diameter) and the chamber a thick one (2 mm. diameter). By using two vessels in this way the counter remained essentially free of contaminant after experimentation since the radioactive content was always small and the fraction sticking to the wall was found to be $\sim$1 per cent. Moreover the thick central wire of the chamber permitted easy attainment of saturating current conditions even at high pressure. On the other hand, lower pressures in the counter facilitated the search for a good plateau for complete counting. The large ratio of length/diameter assured us that end-effect, even if somewhat different for counter and chamber, did not appreciably affect the evaluation of V.

The values obtained for V in the two experiments which seemed to be most satisfactory were as follows :—

Press in cm. of Hg.		V in electron-volts.
CH_4	Argon	
15	135	28·5
10	138	28·4

These two values are not considered to be significantly different and the error limits on the value obtained for X-rays makes it consistent with them. It will be shown below that argon and methane agree closely in V values and we are of the opinion that the V value of the mixture is effectively the same for the β-rays of tritium (average energy 5·7 keV.) and photo-electrons of Cu^{64}. This conclusion is drawn from our earlier curve[7], showing constancy of V as a function of the energy of the electron. We conclude finally then that our work indicates that the value of V for argon containing about 10 per cent CH_4 and for electrons of energy between a few keV. and say 50 keV. is given by

$$V=28{\cdot}5\pm0{\cdot}5 \text{ eV.}$$

Other Gases.

To extend the work to other gases it was only necessary to subidivde a supply of tritium into a number of volumes of accurately known relative values. These were made nearly equal and added to the ion chamber in turn, together with the gas or mixture under investigation. Fairly consistent values of V for several gases were measured in this way but the procedure was not entirely satisfactory, mainly on account of the appreciable difficulty of “ sticking ” of tritium to the walls of the glass containers. However, the value for air given below was obtained by this method, the standard reference gas being argon. Taking the value of V for argon as 28·5 eV. the value for air was by comparison 31·0 eV. Next an ion chamber containing some tritium was filled to atmospheric pressure with nitrogen and the current measured. Oxygen was added to the vessel and the current re-measured. Hence the value for nitrogen was deduced, viz., 32·0 eV. Finally, V for oxygen was found by adding oxygen to give a 50/50 mixture of oxygen and nitrogen, noting the current in the mixture and assuming that the change in the current from the value for nitrogen was proportional to the percentage oxygen content of the mixture. Among the values listed below the oxygen result is the least trustworthy in view of this assumption and the difficulty of measuring accurately the small changes.

Gas	V in eV.
Air	31·0
Nitrogen	32·0
Oxygen	28·8

The value for carbon dioxide was obtained by noting the current passing in a chamber filled to atmospheric pressure with argon. Carbon dioxide was added to raise the pressure to two atmospheres. Great care was taken to measure the true saturation current through the tube in both conditions and it was separately verified that there was no appreciable effect which could be ascribed to mere pressure increase. Thus at pressures of argon between 30 cm. Hg and 150 cm. Hg no change was noted in the same counter used in a similar fashion. To deduce the value of V for CO_2 it was necessary to assume the fall in current produced by adding CO_2 to argon was proportional to the amount added. In this way the value V=33·5 eV. was measured for CO_2.

Methane and Ethylene.

The determination of V for these organic gases was done in more detail. The current due to some tritium in pure argon at one atmosphere was measured, then the current was measured as increasing amounts of methane were added till a total pressure of two atmospheres was reached. Next, starting with methane at one atmosphere, argon was added in increasing amounts. It was found that throughout the range pure methane to pure argon there was no appreciable alteration in the ion current. We can therefore deduce that methane has the same V-value as argon, namely 28·5 eV. and that this value holds for any mixture of argon and methane.

Similar examination of the behaviour of ethylene and argon gave the value V=30·6 eV. for ethylene and indicated that the addition of C_2H_4 to argon raised V from 28·5 eV. to 30·6 eV. linearly with increasing percentage of ethylene.

The complete list of values for the gases investigates is thus as below :—

Gas	V in electron–volts
Argon	28·5
Air	31·0
Nitrogen	32·0
Oxygen	28·8
Carbone dioxide	33·5
Methane	28·5
Ethylene	30·6

Discussion.

With regard to the results listed in the table we have to note that no special effort was made to purify the gases thoroughly. Indeed in all cases the values were obtained with a small amount of tritium and hydrogen present (at partial pressures $\gtrsim 1$ cm. Hg). The gases were of commercial purity which meant that there was ~1 per cent nitrogen in the argon and ~2 per cent in the methane. Other contaminations were considerably less in quantity. There was some indication that the value for very pure argon could differ appreciably from 28·5 eV. but no serious attempt was made to establish this observation.

Concerning air, the experiments do not support the variation of V with energy as given by the formula of Gerbes (1935). The value of 31·0 eV. is indeed remarkably close to his value of 31·6 eV. for high speed electrons (60 keV.) and would seem to suggest that V is almost constant for air throughout the range from ~3 keV. to ~300 keV. since the measured value for fast β-rays (energy=·3 MeV.) is 32·0 eV. The best mean value in this large range would appear to be 31·5 eV. It is hoped that future work will improve the accuracy of the value given here. The same method can be extended to include the nearly homogeneous photo-electrons of A^{37} (energy 2·8 keV.) and thus to examine more thoroughly the question of possible variation of V with energy.

We have to note some measure of disagreement with the result of Nicodemus (1946) for electrons of energy 17·4 keV. in argon. His value of 26·9 eV. and our value of 28·5 eV. give a mean value of about 27·7 eV. which is probably just outside the estimated limits of error (believed to be about 0·5 eV.). (We are indebted to Dr. Nicodemus for private communication on his measurements.) It is possible that there is some slight variation of V with electron energy in the range 5·7 keV. (tritium) to 17·4 keV. but this would appear unlikely in view of previous work by the authors and by Hanna, Kirkwood and Pontecorvo (1949). It is hoped that further more extensive measurements will resolve this difficulty.

Regarding some comparable measurements with α-rays (and protons) it appears that the above values for electrons are smaller (Gray 1944) in the case of air (and presumably nitrogen), but larger for ethylene and argon (Schmeider 1939). The values for carbon dioxide are nearly the same. Apparently there is no change of a regular nature involved here.

Further Remarks on Apparatus.

The absolute value of current through the ion chambers was determined by measuring a high resistance in terms of a standard one of 1 megohm. The actual resistance used in the grid circuit of the electrometer triode was $1{\cdot}8\times10^{11}$ ohms. Currents of about 10^{-12} amps. were passed through this and the anode current of the triode was measured with a sensitive micro-ammeter. In practice the standing anode current was cancelled and the meter was reset to zero during ion current measurement by applying a biassing voltage. This bias voltage was the actually observed quantity and it was measured on a standardized voltmeter.

In the course of the work various forms of counters and ion chambers were tried in a search for the most stable design. The later more accurate work was done with ion chambers constructed entirely of glass except for the central wire and its grounding shield. The glass envelopes were coated with aquadag externally and this coated glass was found to serve admirably as cathode. Even pyrex glass of specific resistance 5×10^{16} ohms could be used (at least in thicknesses up to 1·3 mm.). Long term operation brought polarization effects into evidence but these were not serious and the effective capacity of the chamber was such that in short-term operation ($\sim$1 hr.) it was almost negligible. Excellent saturation curves were obtained. Further information (Cockroft and Valentine) will be published separately. Meanwhile it is to be noted that such vessels are well adapted for outgassing in work with gases of high purity or with gases or vapours which tend to be absorbed by the metals in common use.

Further work.

The method of using slow β-sources such as H^3 is well suited to the investigation of the energy expendure in nearly all gases and vapours ; the measurements will be extended soon to cover some specially interesting ones such as hydrogen, helium and water vapour. Knowledge of V for the first two would assist in deriving a theoretical explanation of the results. Moreover any variation of V with energy of the ionizing electrons can be readily investigated with the proportional counter since the spectra of pulses produced by various sources of β-radiation can be analysed and the average energy of the β-particles deduced.

We thank Professor P. I. Dee, F.R.S., for his advice and interest in the work described here.

REFERENÇES.

COMPTON, A. H., and ALLISON, S. K., 1936, *X-rays.*
GRAY, L. H., 1944, *Proc. Camb. Phil. Soc.*, **40**, 72.
BINKS, W., 1936, *Reports on Progress in Physics*, **3**, 347.
EISL, A., 1929, *Ann. Phys. Leipzig.*, **3**, 277.
GERBES, W., 1935, *Ann. Phys.*, **23**, 648.
NICODEMUS, D. B., 1946, *Thesis*, Stanford University.
CURRAN, S. C., COCKROFT, A. L., and ANGUS, J., 1949, *Phil. Mag.*, **40**, 929.
CURRAN, S. C., ANGUS, J., and COCKROFT, A. L., 1949, *Phil. Mag.*, **40**, 36.
CURRAN, S. C., ANGUS, J., and COCKROFT, A. L., 1949, *Phys. Rev.*, **76**, 853.
SLACK, L., OWEN, G. E., and PRIMAKOFF, H., 1949, *Phys. Rev.*, **75**, 1448.
JENKS, G. H., GHORMLEY, J. A., and SWEETON. F. H., 1949, *Phys. Rev.*, **75**, 701.
HANNA, G. C., KIRKWOOD, D. H. W., and PONTECORVO, B., 1949, *Phys. Rev.*, **75**, 985.
SCHMËIDER, K., 1939, *Ann. d. Phys.*, **35**, 445.

XLVIII. *Magnetic Spectrometer of Large Solid Angle.*

By E. R. Rae,
Department of Natural Philosophy, University of Glasgow *.

[Received February 7, 1950.]

[Plate XIII.]

SUMMARY.

An account is given of a magnetic spectrometer of large solid angle and good resolution which should find applications in the study of complex beta-ray spectra by the method of beta–gamma coincidences. The original apparatus, consisting of a segmented annular G.M. counter set round a radioactive source in a magnetic field is described, the method of use being to select from the output of the counter, large pulses which correspond to the simultaneous triggering of several sections by the passage of an electron tangentially to the annular counter. The performance of the instrument is discussed and modifications are described which rectify the defects in the earlier design and give a solid angle of 2·5 per cent of 4π with a resolution of 5 per cent.

The geometry of this type of spectrometer is considered and the performance which might be obtained with good design is discussed. It is concluded that collecting angles up to the order of 50 per cent of 4π are possible with good resolution.

§ 1. INTRODUCTION.

AN account has recently been published by Siegbahn and Slätis (1949) of an instrument for the determination of the energies of beta and gamma radiations. The device consists of an arrangement for placing the source of the radiations at the centre of an annular G.M. counter whose plane is at right angles to the direction of the magnetic field. Considerable work has been done here on an apparatus which is superficially very similar to that of the above authors, but which in fact has several advantages over it. For this reason it has been decided to place on record this short account of our method which confirms some of the results of Siegbhan and Slätis and which also leads to the design of a magnetic spectrometer of very large solid angle which should prove a valuable tool in the examination of complex beta-ray spectra by the method of beta–gamma coincidences.

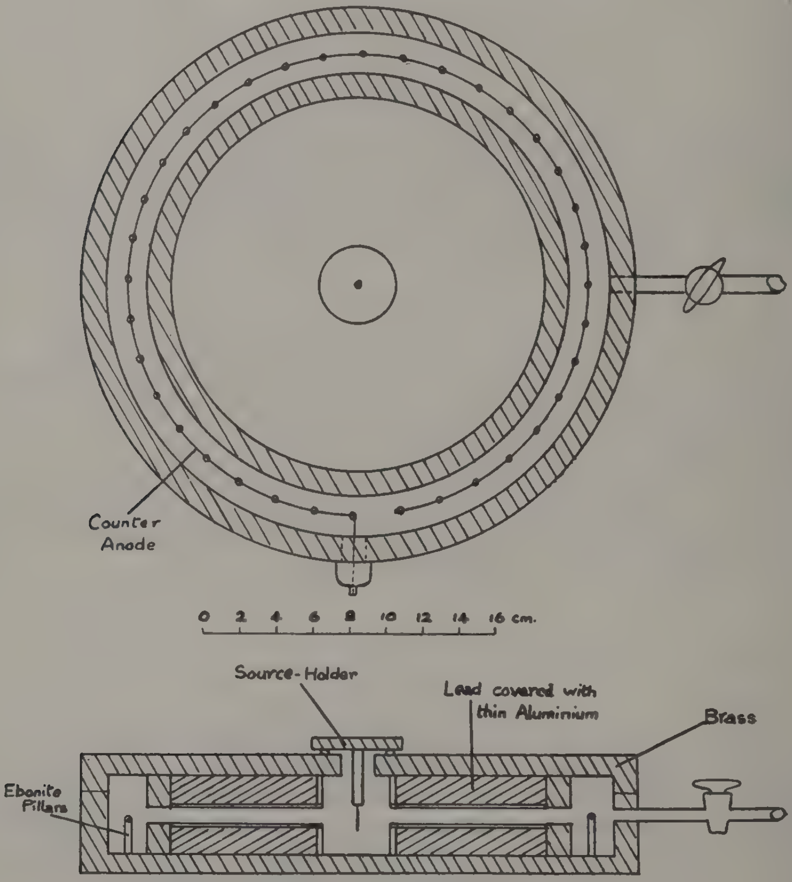
---

* Communicated by the Author.

§2. DESCRIPTION OF ORIGINAL APPARATUS.

The device as originally designed by Dr. S. C. Curran consisted of an annular counter with the wire supported on a large number (about 40) of small insulating pillars which had the effect of splitting the main counter into a large number of small counters acting in parallel (Stever 1942, Curran and Rae 1947). A sketch of the apparatus is given in fig. 1 and it will be seen that the counter had no window, the whole apparatus being filled with the counting mixture. The intention was to use the device to measure the energy of hard gamma-rays by detecting the positron–electron pairs ejected from a thin lead foil placed round the source, for in the case of equipartition of energy, and for the correct value of the magnetic field, both positron and electron could trigger four or five sections of the counter, thus giving a pulse eight to ten times as large as the background single section pulses.

Fig. 1.

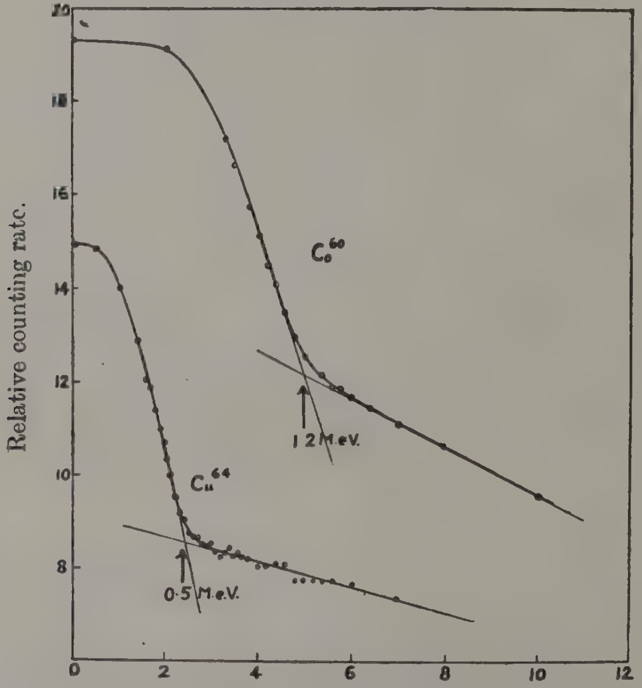


Schematic diagram of original apparatus.

Unfortunately this simple and elegant arrangement did not yield the expected results, for two main reasons: firstly because the background was found in fact to contain a large number of double and multiple pulses owing to bad geometry which made it highly probable that a particle entering one section of the counter would be scattered into at least one more. This meant that there was a considerable number of four and five-fold pulses when a strong gamma-ray source was being used, and so the number of accidental coincidences between these, giving eight to ten-fold pulses was not negligible. The second difficulty was that because the counter sections were so short relative to their length, they had very serious end effects which caused them to have only a very short plateau. This meant that even on the plateau the pulse size was a steep function of the high tension voltage, and so it was difficult to get stable working and reproducible results, since the method depended on selecting pulses of a particular size with a discriminator. These same two difficulties also prevented the device from being used satisfactorily in a similar manner to determine ordinary beta-ray and secondary electron spectra. It was at this point that the method of removing the discriminator and counting all the pulses of all sizes was tried, since though it does not give the spectrum of the source directly, it does overcome the difficulties mentioned above. This is of course the same method as that used by Siegbahn and Slätis, and though it was not studied here in detail, the results obtained were very similar to those published by them. For comparison the curves obtained for the secondary electrons ejected from a thick lead foil by the gamma-rays of Co^{60} and the annihilation radiation of Cu^{64} are shown in fig. 2. From these results and a number of others, an empirical calibration curve was constructed for the device as used in this way.

This method is however subject to several serious disadvantages. Firstly, even when one has obtained the experimental integral spectrum there is no explicit way of obtaining the true spectrum of the source because of the fact that for any setting of the magnetic field the solid angle of the device is a function of the energy of the particles detected. Thus when used in this way, the device is only suitable for determining end-points, unless one is prepared to carry out long and tedious successive approximations. Secondly, because of the high background in the counter, photo-electron lines and any other kind of fine structure are quite invisible, and thick convertors must be used for the determination of gamma-ray energies, which makes the resolution of the device poor. For instance, it will be noted that in the curves for Co^{60} given by Siegbahn and Slätis and by the present author, in neither case was it possible to resolve the lower energy line at 1·1 MeV. from that at 1·3 MeV. Thirdly, as stressed by Siegbahn and Slätis, the effect of scattering in the gas is such as to produce counts at values of the magnetic field much higher than that corresponding to the true end-point of the spectrum being examined, and this causes difficulties in the interpretation of the curves. In addition to these criticisms it should be mentioned that since one of the obvious uses for a large solid angle spectrometer of this sort would be the separation of complex beta-ray spectra by the method of beta–gamma coincidences, it would be a great advantage if the device could be made to give the true spectrum instead of an integral one, since the former would reduce the background of accidental coincidences by a very large factor near the end-point, that is, in the region of the greatest interest. For these reasons it was decided that it was worth while to attempt to overcome the difficulties in the original design in a different way.

Fig. 2.



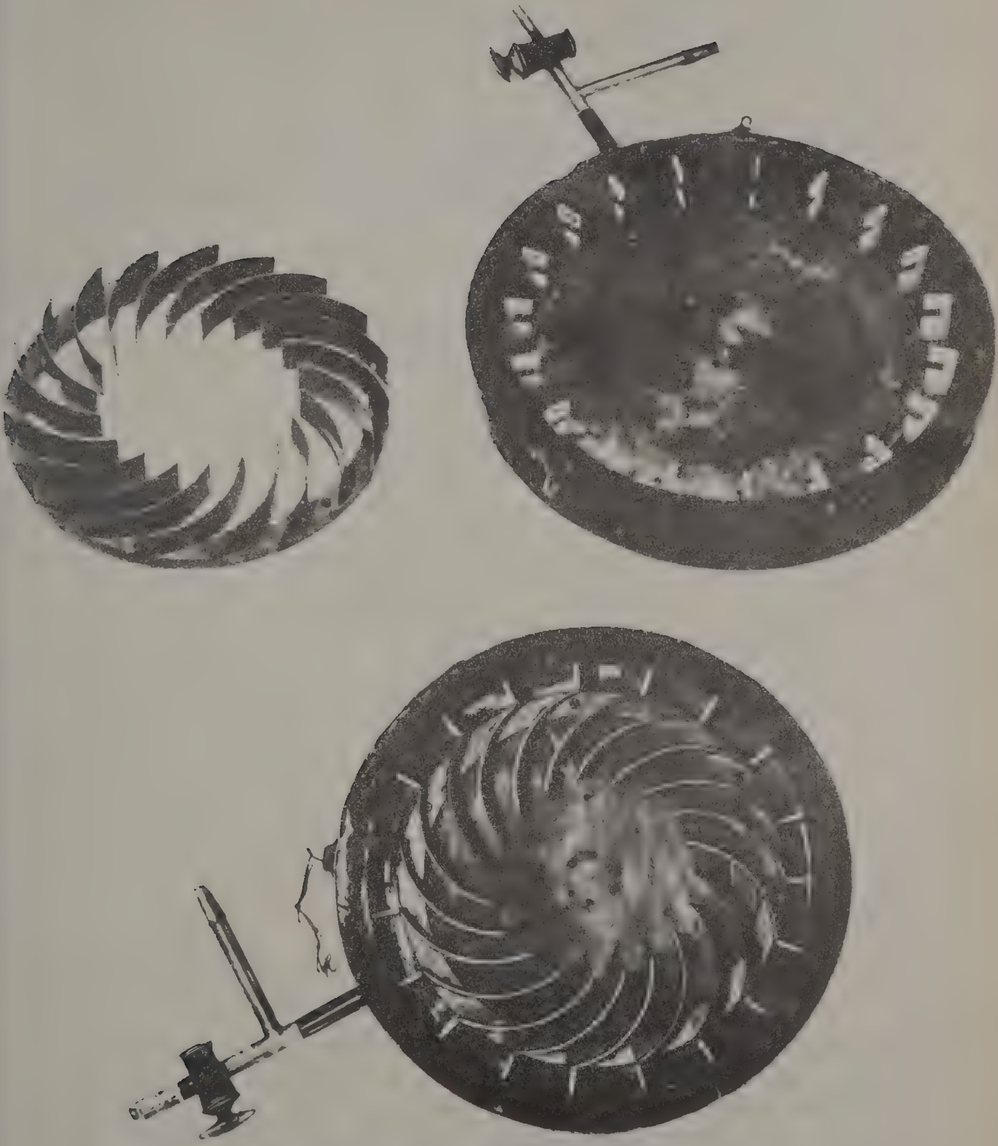
Integral momentum spectra of electrons ejected from a thick lead foil by the gamma-rays of Co^{60} and the annihilation radiation of Cu^{64}.

§ 3. THE IMPROVED APPARATUS.

A photograph of the improved apparatus is shown in fig. 3 (Pl. XIII.) and it will be seen that the number of sections into which the counter was divided has been reduced, thus giving longer sections and very much improved electrical performance. It will also be noted that the sections are separated by polythene partitions instead of the pillars of the original design, and that these partitions have openings in them just large enough to permit the passage of electrons within the desired momentum range.

RAE

Fig. 3.



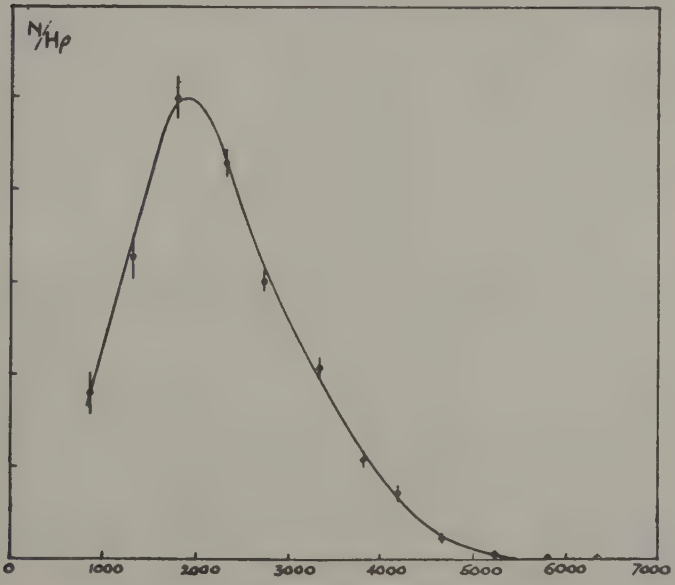
Photograph of improved apparatus.

To face page 528

This has the effect of cutting down to a very large extent the number of particles which are accidentally scattered into more than one section. Finally, the device has been fitted with a set of baffles which permit only particles of the correct sign of charge and the correct momentum range to enter the counting system, and which together with the partitions ensure that only those particles are detected which have not been seriously scattered in the gas.

With these improvements it was found possible to take spectra with the discriminator set to count single, double, or triple section pulses with good reproducibility. The single section spectra contained a large number of spurious counts at low energies because of scattering, but the spectra obtained from the double and triple section pulses were very similar. The spectrum of RaE as obtained from an old thin-walled radon seed with triple section pulses is shown in fig. 4 and it will be observed that the shape is good apart from a displacement of the number maximum towards higher energies because of the scattering of the softer beta-rays in the gas filling. The filling for this run was 3 cm. argon+1 cm. alcohol, the working voltage about 1100 V. and the estimated resolution (for triples) of the order of 5 per cent.

Fig. 4.



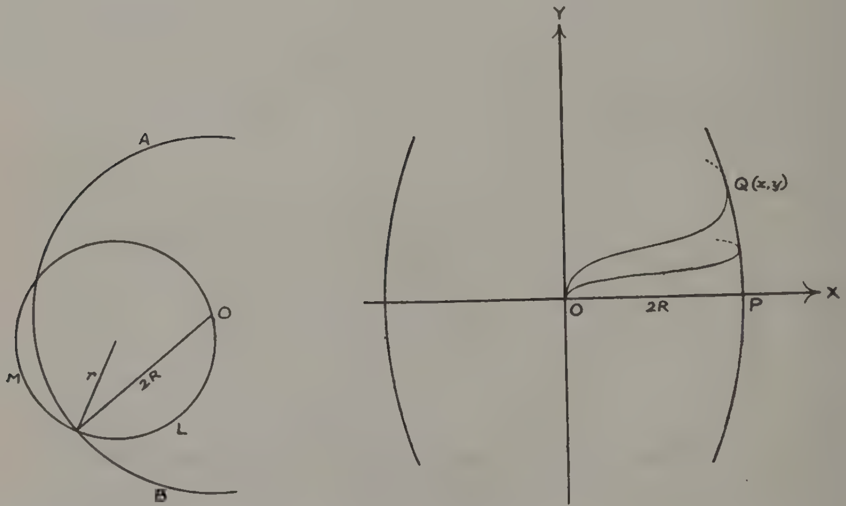
Hρ in Gauss. Cm.

The beta-ray spectrum of an old radon seed.

These modifications to Curran's original apparatus thus produced a 360° spectrometer with a resolution of about 5 per cent and a solid angle of 2·5 per cent of 4π as compared with 1 per cent of 4π for a corresponding semicircular focusing spectrometer. It can however easily be shown that by altering the type of counting system and so obtaining a much wider slit, an instrument can be built which will give a solid angle of 20 per cent of 4π with a 5 per cent geometrical resolution, and further, that if the apertures in the partitions between the counters and the corresponding baffle systems are specially shaped, then solid angles up to the order of 50 per cent of 4π might be obtained with 5 per cent resolution. The design considerations for such an instrument both in the restricted and general form are discussed in the next section.

§ 4. Design Considerations for Spectrometer of very large Solid Angle.

Fig. 5.



(*a*) Section by median plane.

(*b*) Section by plane containing the axis of symmetry.

Geometry of 360° spectrometer.

If we consider the diagram in fig. 5 (*a*) which shows a section of the spectrometer by the median plane, then the circle AB represents the inner edge of the counting area which has radius 2R. OLM represents the path of a particle, radius r, where r is just great enough to ensure that it intersects n counter sections. Now it can easily be shown that if α is the angle subtended at O by a single section, then

$$r/\mathrm{R}=\sec\,(\overline{n-1}\,\alpha/2).$$

Hence we can draw up a table of values of r/R as a function of n and α, which table will tell us how wide we must make the counting area and what order of pulse size we must record to obtain a given resolution with a given value of α. With $\alpha=15°$, which is a suitable value for ease in construction, we find that the values of r/R for $n=1, 2, 3, 4, 5$ are 1·000, 1·009, 1·034, 1·082 and 1·154. This means that if we set the limit of our counting area at $2R_1=1{\cdot}07\times 2R$ and count triples, then we have a resolution of the order of 5 per cent.

If we now consider fig. 5 (*b*) which represents a section of the spectrometer by a plane containing the axis of symmetry, then the magnetic field lies along OY and O represents, as in fig. 5 (*a*), the position of the source which in this case is assumed to be emitting mono-energetic particles. Then OP represents the projection in the plane of the paper of the semicircular trajectory of a particle emitted normally to the plan of the paper. The sine curve OQ represents the projection on the plane of the paper of the helical trajectory of a particle emitted in a plane through O perpendicular to OP but whose direction of emission makes and angle θ with the normal to the plane of the paper. Then if Q has coordinates (x, y) with respect to O as origin, it is clear that if $OP=2R$, the locus of Q is the ellipse given by the freedom equations $x=2R \, . \cos\theta$, $y=\pi R \, . \sin\theta$, *i.e.* the ellipse

$$x^2/(2R)^2+y^2/(\pi R)^2=1.$$

This means that the locus of the points of maximum distance from the axis of symmetry, of the particles emitted from O, is the ellipsoid of revolution :—

$$(x^2+z^2)/(2R)^2+y^2/(\pi R)^2=1.$$

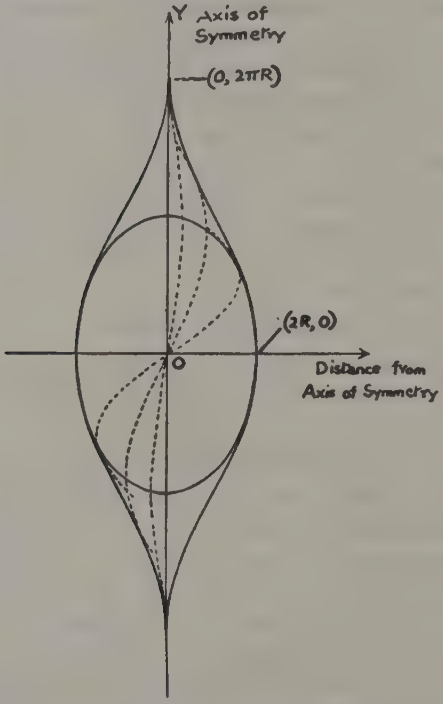
Now it can easily be shown that if we consider only those particles which are emitted with angles θ lying between $\theta=\pm\phi$ then the solid angle into which they are projected is $4\pi\sin\phi$. Hence for a geometry to collect 10 per cent of all the particles from the source $\sin\phi=1/10$, and for a limiting particle $x/2R=\cos\phi=1-1/200$. That is, the substitution of a right circular cylinder of radius 2R for the ellipsoid would introduce an error of only ½ per cent in the momentum of the particles registered. For 20 per cent collection the error would be about 2 per cent and for larger solid angles would increase rapidly. Thus where an overall resolution of the order of 5 per cent is desired, a spectrometer of solid angle 20 per cent of 4π could be built on the same principle as that described in the last section with simple cylindrical construction and rectangular apertures in the partitions. For larger solid angles, the partitions would have to be shaped, the shape being given by the envelope of the family of sine curves corresponding to the plot of the actual distance of the particle from the axis of symmetry against y. The general shape of this curve is shown in fig. 6 and it has freedom equations given by

$$x^2=-4R^2\sin^3X/\cos X(X-\tan X),$$
$$y^2=\ \ 4R^2X^3/(X-\tan X),$$

where X is a parameter whose values are confined to those that make tan X/X negative.

It is clear, however, that for these large solid angles the original scheme for having a common anode for all the counters running round the median plane would have to be abandoned since the depth of the counting space has become so great. The difficulty can be overcome by using counters of roughly square section when viewed along the axis of symmetry, and fitting separate anodes parallel to the magnetic field.

Fig. 6.



Section by plane through axis of symmetry, of envelope of trajectories.

Even with this method of construction, however, a reference to fig. 6 will show that solid angles in excess of about 50 per cent of 4π would produce counters of such a shape that their efficiency might be rather poor for particles intersecting them near their ends. For this reason and because of the difficulty of constructing a baffle system or thin window of the correct shape, it seems probable that 50 per cent collection would be about the practical maximum. It should of course be realized that these figures represent the ideal collecting fractions which might be obtained with perfect design; in practice any baffle system would obstruct part of the solid angle and any windows would cause the loss of particles by scattering as would also the gas filling. Nevertheless, the figures given above represent such an improvement on the collecting fractions available by any of the conventional methods of beta-ray spectroscopy that it is felt that the development of this method would provide a powerful tool in the study of complex beta-ray spectra by the method of beta–gamma coincidences.

I must acknowledge the help given me in this work by Dr. S. C. Curran who designed the original apparatus and with whom I had many dicussions which were of the greatest value in suggesting ways of overcoming the faults in the early design. I must also thank Mr. D. L. Pursey for deriving the freedom equations to the envelope of the electron trajectories.

REFERENCES.

CURRAN, S. C. and RAE, E. R., 1947, *J. Sci. Instrum.*, **24**, No. 9, 233.
SIEGBAHN, K. and SLÄTIS, H., 1949, *Arkiv for Mat., Astr. o. Fysik*, **36** A, No. 22, p. 1.
STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.

XLIX. *The Influence of Exchange and Correlation Forces on the Specific Heat of Free Electrons in Metals.*

By E. P. WOHLFARTH, Ph.D.,
Department of Mathematics, Imperial College, London*.

[Received March 1, 1950.]

SUMMARY.

The effect on specific heat of including exchange terms in the expression for the electronic energy of a free electron gas is discussed. Previous work is considered; the only full discussion (Koppe 1947) contains a number of algebraical errors. When these are corrected the theoretical expression for specific heat becomes, for sodium,

$$C_e=\gamma T/(8{\cdot}9-1{\cdot}7\log_{10}T),$$

over a wide temperature range, γ being the Sommerfeld factor (§ 2). This relation is in disagreement with observation, due to the approximation implied in the use of a one-electron wave function. In § 3 Landsberg's tentative suggestion (1949) for estimating the effects of correlation forces is considered in relation to the present problem, a screened rather than a simple Coulomb potential being used in the exchange integral. The value of the screening constant required to restore approximate agreement with observed specific heats is found to be of the same order of magnitude as that derived from the soft X-ray emission spectrum.

§ 1. INTRODUCTION.

THE well known Sommerfeld treatment of the electronic specific heat of metals leads to a linear temperature dependence at low temperatures. This result agrees with experiment not only for metals for which the operative electrons may be assumed very nearly free (*e. g.* sodium and copper) but also for the transition metals (*e. g.* nickel and palladium), the specific heat of which is due largely to the holes in an incompletely filled energy band (*cf.* Jones and Mott 1937). In the former case the observed linear temperature coefficient usually differs only slightly from Sommerfeld's value (for copper the ratio is about 1·5); such discrepancies may be justifiably ascribed to slight deviations of the electron distribution from that of perfectly free electrons. For the transition metals the ratio is much larger (about 30 for palladium), consistent with the fact that the density of states near the Fermi limit in the incompletely filled band is very high. In all cases, however, the observed temperature dependence is strictly linear.

---

* Communicated by the Author.

A closer examination of the electron theory of metals reveals at once, however, that the simple treatment is incomplete, owing to the neglect of the exchange terms, which have to be included in the expression for the electronic energy as a consequence of the exclusion principle. This omission is inconsistent with the application of theory to ferromagnetism (*cf.* Wohlfarth 1949) and to the calculation of cohesive energies (*cf.* Seitz 1940), where the exchange terms play a vitally important part. If the treatment is extended to include the exchange terms it is found that the agreement between calculated and observed specific heats is to a large extent invalidated. This situation is discussed in the present paper. It has been discussed previously, and a brief account of the history of this aspect of electron theory may be of interest.

Perhaps the first mention of the anomaly was made by Brillouin (*e. g.* 1934), who showed that on applying to metals the methods of Fock and Dirac, in which exchange is taken into account, the dependence of the energy on momentum is modified considerably as compared with that given by Sommerfeld, in that the $\epsilon(k)$ curve has an infinite slope, so that the density of states vanishes, at the top of the Fermi distribution. The effect of this modification on specific heat was first mentioned by Bardeen (1936), who found that the specific heat of free electrons is given, as a result of including the exchange terms, by

$$C_e = \gamma T / |\ln T| \qquad \text{(1.1)}$$

at very low temperatures, instead of the usual linear relation. At higher temperatures the specific heat is stated to be smaller than the Sommerfeld value by a large factor of about 3–7. Bardeen never fully published his findings (presented in an Abstract of some 300 words), and in a later paper by Wigner (1938) little more is done than to give a *resumé* of the essential ideas and to quote an explicit relation for the ratio of the Sommerfeld specific heat to that calculated by Bardeen (*cf.* § 2). The first *full* discussion of the problem was given by Koppe (1947). Unfortunately Koppe's paper contains a number of algebraical errors which almost entirely destroy the significance of his final conclusions. Although obtaining relation (1.1) at very low temperatures, Koppe states that it is valid only below 10^{-4} °K. and that at higher temperatures $C_e = \gamma T$, with γ only slightly less than the Sommerfeld value, the calculated ratio being 1·58 for sodium at 1° K. as compared with the much larger value obtained by Bardeen.

In the present paper Koppe's calculations are corrected and the amended results briefly discussed (§ 2). In § 3 a possible method of circumventing these difficulties is considered, based essentially on a suggestion due to Landsberg (1949). Briefly, the suggestion concerns the possibility of estimating by a tentative empirical method the effects on the properties of the electron gas of the correlation energy terms (arising from the fact that the initial total wave function contains terms depending on the inter-electronic distances and is thus more complicated than that based on the usual one-electron approximation). That these correlation terms might restore the agreement with experiment was in fact first suggested by Wigner (1938).

§ 2. CORRECTION OF KOPPE'S CALCULATIONS.

Koppe derives an expression (1947, (2)) for the free energy of an assembly of free electrons, including the exchange energy. This is derived in the ususal way (*cf.* Sommerfeld and Bethe 1933); in considering thermodynamic properties of the assembly at finite temperatures it is, however, necessary to introduce the Fermi distribution function $f(k, T)$ and the free energy expression becomes

$$\begin{aligned} F(T) = \frac{h^2 V}{16\pi^4 m}\int_0^\infty f(k, T)k^4\, dk \\ - \frac{e^2 V}{4\pi^3}\int_0^\infty\int_0^\infty f(k_1, T)f(k_2, T)k_1 k_2 \ln\frac{k_1+k_2}{|k_1-k_2|}\, dk_1\, dk_2 \\ +\kappa T\frac{V}{2\pi^2}\int_0^\infty \{f\ln f+(1-f)\ln(1-f)\}k^2\, dk, \end{aligned} \quad . \; . \; . \; . \quad (2.1)$$

the second term representing the exchange energy, k denoting wave number (related to Koppe's p by $p=kh/2\pi$) and V the volume of the assembly. By minimizing F subject to constancy of the total number of particles, Koppe finds

$$\frac{h^2}{8\pi^2 m}k_1^2 - \frac{e^2}{\pi k_1}\int_0^\infty k_2 f(k_2, T)\ln\frac{k_1+k_2}{|k_1-k_2|}\, dk_2 - \zeta = \kappa T\ln\frac{1-f(k_1, T)}{f(k_1, T)}, \quad . \quad (2.2)$$

ζ being the usual statistical parameter. This is a functional equation of considerable complexity, and Koppe approximates to the solution by the discontinuous function

$$\left.\begin{array}{ll} f(k, T)=1, & 0\leqslant k\leqslant a \\ f(k, T)=\alpha-\beta k, & a\leqslant k\leqslant b \\ f(k, T)=0, & b\leqslant k \end{array}\right\}, \quad . \; . \; . \; . \; . \quad (2.3)$$

with

$$a=(\alpha-1)/\beta, \qquad b=\alpha/\beta.$$

The parameters α and β are eliminated through the relation for the total number of particles (if each state is occupied by two particles with opposite spin,

$$N=\frac{V}{\pi^2}\int_0^\infty k^2 f(k, T)\, dk), \quad . \; . \; . \; . \; . \; . \; . \quad (2.4)$$

and by minimizing F. For $T=0$, $\alpha\to\infty$ and Koppe evaluates the various terms in (2.1) in series form in α^{-1} to cover the low temperature range, neglecting α^{-3} and higher powers. The subsequent development shows that the trial function (2.3) probably represents a very close approximation to the solution of (2.2).

Koppe's main error arises in evaluating the double integral in (2.1); on carrying through the calculation correctly this is found to be

$$\frac{1}{2}\left(\frac{\alpha}{\beta}\right)^4\frac{e^2V}{4\pi^3}\left\{1-2\alpha^{-1}+\frac{119}{72}\alpha^{-2}-\frac{1}{6}\alpha^{-2}\ln 2\alpha+O(\alpha^{-3})\right\}, \quad . \quad (2.5)$$

instead of Koppe's expression (6, 2). Using (2.5) and eliminating (α/β) through (2.4), the expression for F becomes

$$F(T)/N=\frac{3}{5}\epsilon_0\left(1+\frac{5}{12}x^2\right)-\frac{3}{4}\epsilon_j\left(1-\frac{13}{72}x^2+\frac{1}{6}x^2\ln\frac{x}{2}\right)-\frac{3}{2}\kappa Tx\ldots, \quad (2.6)$$

with $x=\alpha^{-1}$ and

$$\epsilon_0=(h^2/8m)(3N/\pi V)^{2/3}, \qquad \epsilon_j=e^2(3N/\pi V)^{1/3}.$$

Koppe's corresponding quantities (his (9,1) and (9,2)) are related to ϵ_0 and ϵ_j by

$$\theta=3\kappa T/\epsilon_0, \qquad q=3\epsilon_j/2\epsilon_0\,;$$

in employing them he makes a further error in that his relation (10) should read

$$\theta=\left(1+\frac{97}{3}q\right)x-\frac{5}{3}qx\ln\frac{x}{2},$$

using his (incorrect) relation (7) for F. The correct relation between x and T is obtained through $(\partial F/\partial x)_T=0$, giving, from (2.6),

$$3\kappa T/\epsilon_0=\left(1+\frac{7}{24}\frac{\epsilon_j}{\epsilon_0}\right)x-\frac{1}{2}\frac{\epsilon_j}{\epsilon_0}x\ln\frac{x}{2}, \quad . \quad . \quad . \quad . \quad (2.7)$$

and the specific heat expression becomes

$$C_e=-T\left(\frac{\partial^2F}{\partial T^2}\right)_V=\frac{3}{2}N\kappa T\frac{dx}{dT}=\frac{9/2\;N\kappa(\kappa T/\epsilon_0)}{d(3\kappa T/\epsilon_0)/dx}$$

$$=C_e^0\Big/\left\{1-\frac{5}{24}\frac{\epsilon_j}{\epsilon_0}-\frac{1}{2}\frac{\epsilon_j}{\epsilon_0}\ln\frac{x}{2}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.8)$$

where $C_e^0=9/2\;N\kappa(\kappa T/\epsilon_0)$, the value obtained for specific heat in the absence of exchange, *i. e.* the Sommerfeld value. The factor 9 appears instead of the usual π^2 owing to the approximation (2.3).

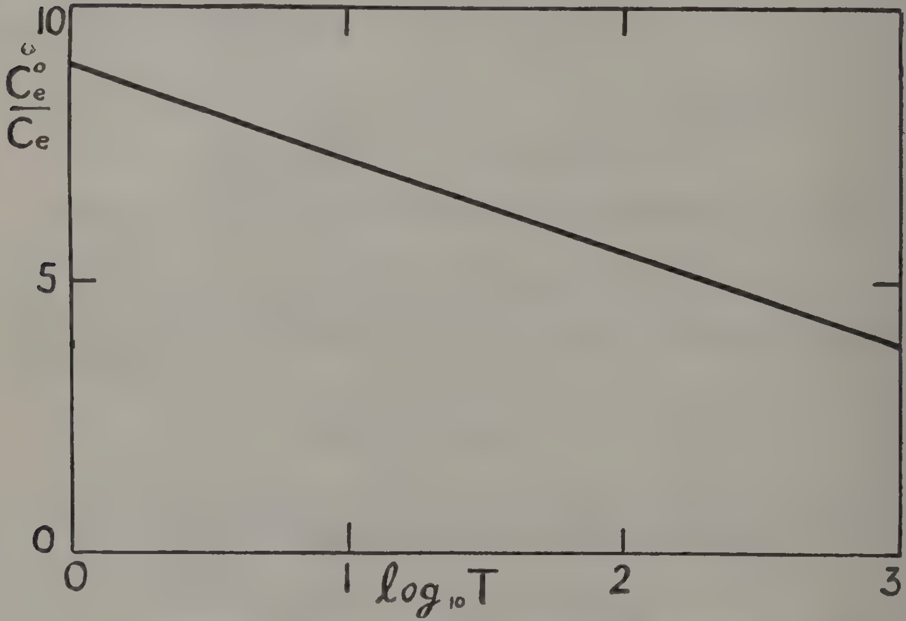
For sodium $\epsilon_0/\kappa=3{\cdot}66\times10^4$ °K., $\epsilon_j/\kappa=4{\cdot}84\times10^4$ °K. so that (2.7) and (2.8) may be rewritten as

$$\left.\begin{aligned} C_e^0/C_e&=0{\cdot}72-1{\cdot}52\log_{10}y,\\ 4{\cdot}10\times10^{-5}T&=1{\cdot}38y-1{\cdot}52y\log_{10}y,\end{aligned}\right\}\quad . \; . \; . \; . \; . \quad (2.9)$$

with $y=x/2=(2\alpha)^{-1}$. The solution of (2·9) has been obtained graphically and it is found that for $\log_{10} T$ between 0 and 3, and probably for even higher temperatures, as long as $T\ll\epsilon_0/\kappa$ the electronic heat of sodium is given by

$$C_e^0/C_e=8{\cdot}9-1{\cdot}7\log_{10}T. \quad . \; . \; . \; . \; . \; . \; . \quad (2.10)$$

Fig. 1.



Decrease of electronic heat due to exchange, for sodium.
C_e, electronic heat with exchange ; $C_e^0=\gamma T$, Sommerfeld electronic heat.
Cf. relation (2·10).

This relation may be compared with that due to Bardeen, quoted by Wigner (1938) and stated to be "sufficiently accurate for temperatures between 1 and 100° K". Bardeen obtains

$$C_e^0/C_e=1+\frac{4\pi me^2}{h^2k_0}\ln\frac{2e^2k_0}{\pi\kappa T}, \quad . \; . \; . \; . \; . \; . \quad (2.11)$$

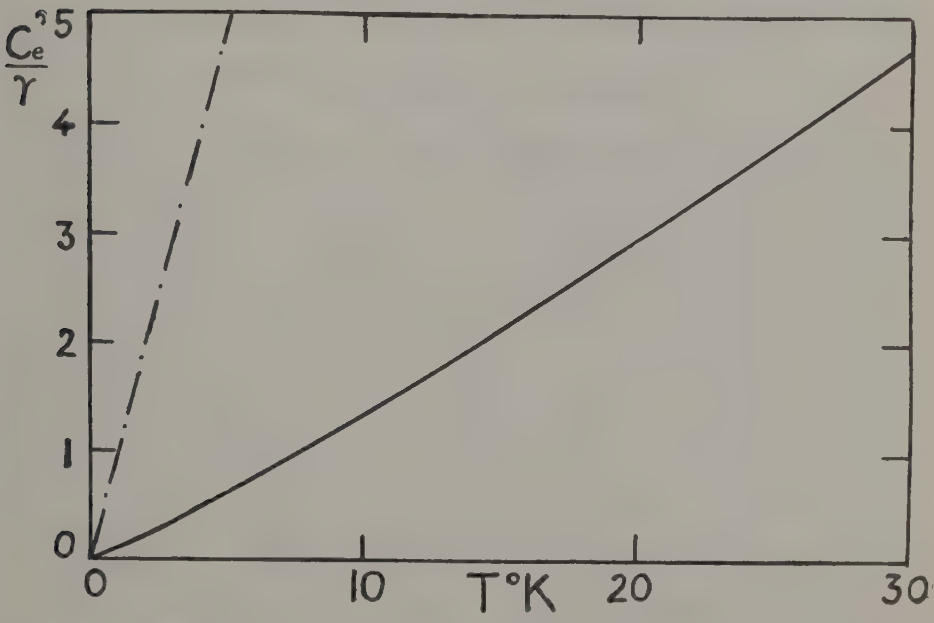
k_0 being the radius of the Fermi sphere. For sodium $k_0=0{\cdot}909\times10^8$ cm.$^{-1}$ so that (2.11) gives

$$C_e^0/C_e=8{\cdot}7-1{\cdot}5\log_{10}T.$$

It is difficult to see whether the close agreement between these two expressions is more than accidental, since no details as to the derivation of (2.11) are given. (It may be pointed out, however, that (2.11) is easily obtained if certain rough, and apparently unjustifiable assumptions are made.)

Fig. 1 gives a plot of (2.10), and fig. 2 of C_e/γ against T at very low temperatures, with $C_e^0=\gamma T$. The difference between the Sommerfeld value and that obtained by including exchange is seen to be very pronounced, not only at low temperatures where the ratio between them is very large (8·9 at 1° K. and 6·7 at 20° K.) and the C_e-T relation is not linear, but also at higher temperatures where the ratio is still large (3·8 at 1000° K.) although C_e is proportional to T over limited temperature ranges.

Fig. 2.



Variation of electronic heat with temperature, for sodium. Full curve, with exchange; broken curve, without exchange. γ, Sommerfeld factor.

Extensive measurements of low temperature specific heats of metals and alloys have been made. For sodium (Pickard and Simon 1948) an anomalous temperature variation of specific heat between 5 and 10° K. precludes the possibility of estimating the electronic contribution, but for a number of other metals for which the free electron approximation is applicable a linear temperature dependence, with a coefficient of the order of magnitude of the Sommerfeld value, has been established beyond doubt (*cf.* Jones and Mott 1937). For the transition metals where the electronic contribution to specific heat is much higher the same general conclusions have been drawn from some very precise measurements (*e. g.* Pickard and Simon 1948, palladium; Silvidi and Daunt 1950, tungsten); in some cases the linear temperature coefficient has been satisfactorily correlated with other physical properties of the metal. Whereas for the transition metals the effect of the exchange terms on the density of states and the electronic properties is at present incalculable, it is reasonable to suppose that theory would lead to similar anomalous results. It must be concluded, therefore, that the inclusion of the exchange terms in the way described provides an insufficient approximation.

§ 3. APPLICATION OF AN EMPIRICAL METHOD FOR ESTIMATING CORRELATION.

The anomalous results obtained in the preceding section arise as a consequence of the infinite slope at the Fermi surface of the curve relating energy and momentum (Brillouin 1934). Similar difficulties were encountered by Heisenberg (1947) in his theory of superconductivity. Heisenberg overcame this difficulty by considering electrons in a thin spherical shell at the Fermi surface to be excluded from normal electronic processes. More recently Landsberg (1949) found, in attempting a calculation of the width of the " tail " of the soft X-ray emission curve of sodium, in which a similarly divergent result was obtained, that even with reasonable values of the energy width of Heisenberg's shell the calculated tail width is still very much larger than that observed. He concluded that the discrepancy is due to the neglect of correlation forces, as in the present problem. Owing to the difficulties involved in an accurate treatment of correlation (Wigner 1934, 1938) it was suggested that the inter-electronic terms in the energy could be simulated by replacing the Coulomb potential in the exchange integral by a screened potential

$$V(r_{12})=e^2r_{12}^{-1}\exp(-\lambda r_{12}), \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

while the one-electron character of the total wave function is retained. The correlations introduced in this way are effective at large distances and are thus of a different nature from those considered by Wigner, which are effective at smaller distances only. The present method must be regarded as purely empirical, being introduced so as to attain agreement between the results of the theoretical treatment and a variety of experimental results. The use of (3.1) is still unsatisfactory in calculating cohesive energies of metals, since the screening term will affect the Coulomb as well as the exchange energy term in the total energy expression.

On calculating the tail width using (3.1) Landsberg finds that agreement with the observed value is attained if λ is taken to be about $1{\cdot}2\times10^8$ cm.$^{-1}$.

In this section the effect of using (3.1) on the calculated electronic heat is considered. At absolute zero the exchange energy becomes

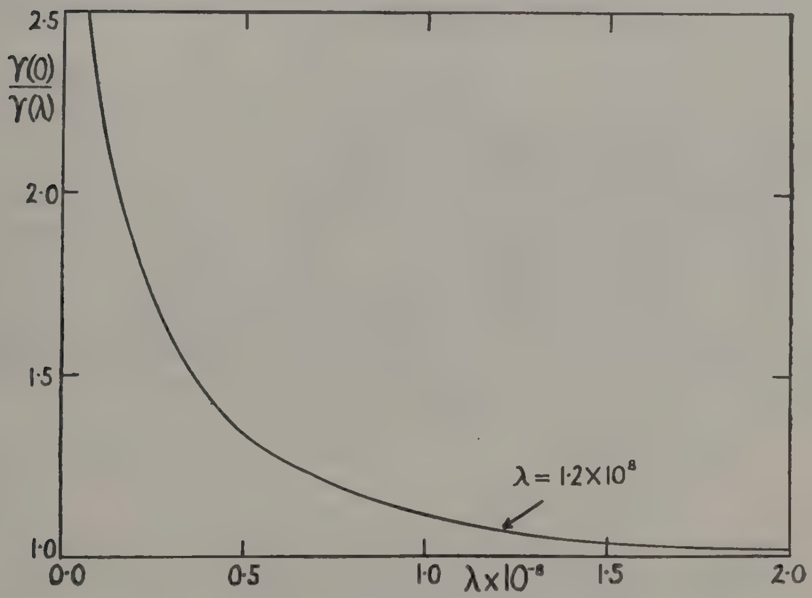
$$J(k)=\frac{1}{8\pi^8 V}\iiint \exp\{i(\mathbf{k}-\mathbf{k}').(\mathbf{r}_1-\mathbf{r}_2)\}V(r_{12})\,d\tau_{12}\,d\tau_{k'}$$

$$=\frac{e^2}{2\pi^2}\int_0^{k_0}\frac{d\tau_{k'}}{|\mathbf{k}-\mathbf{k}'|^2+\lambda^2}$$

$$=\frac{e^2}{2\pi k}\int_0^{k_0}k'\ln\frac{(k+k')^2+\lambda^2}{(k-k')^2+\lambda^2}\,dk'$$

$$=\frac{e^2k_0}{2\pi}\left\{2+\tfrac{1}{2}\left(\frac{1-x^2+l^2}{x}\right)\ln\frac{(1+x)^2+l^2}{(1-x)^2+l^2}-2l\tan^{-1}\frac{2l}{x^2+l^2-1}\right\}, \qquad (3.2)$$

Fig. 3.



Change of electronic heat of sodium due to screening.

λ, screening constant in (3.1) ; γ (0), Sommerfeld linear temperature coefficient ; $\gamma(\lambda)$, coefficient for screened electrons. For $\lambda\neq 0$ it is assumed that $C_e=\gamma(\lambda)T$.

where k_0 is the radius of the Fermi sphere, $x=k/k_0$, $l=\lambda/k_0$. An adaptation of Koppe's procedure to evaluate J at finite temperatres would be difficult, but it is reasonable to suppose that once the infinity in the slope of $J(k)$ at the Fermi surface is removed the specific heat will have a normal temperature dependence at all temperatures except for a residual decrease of γ, depending on λ, below the Sommerfeld value. If this is assumed it is found that

$$\left.\begin{aligned} C_e^0/C_e &= 1+\frac{4\pi m e^2}{h^2 k_0}\, g(l), \\ \text{where} \\ g(l) &= \frac{2+l^2}{4}\ln(1+4l^{-2})-1. \end{aligned}\right\} \quad (3.3)$$

Fig. 3 gives a plot of C_e^0/C_e against λ for sodium, and it is seen that C_e^0/C_e tends rapidly to 1 as λ increases and that for $\lambda = 1{\cdot}2 \times 10^8$ cm.$^{-1}$ it is only about 8 per cent above unity. It may be shown that for this value of λ the density of states curve has very nearly normal form so that the above assumption is justified.

It appears, therefore, that using (3.1) in calculating specific heats the agreement with the simple theory may be regained, and that the value of λ required is of the same order as that deduced to bring about agreement with the observed soft X-ray emission spectrum.

REFERENCES.

BARDEEN, J., 1936, *Phys. Rev.*, **50**, 1098.
BRILLOUIN, L., 1934, *Helv. Phys. Acta*, **7** (Suppl. 2), 33.
HEISENBERG, W., 1947, *Z. Naturforsch.*, **2a**, 185.
JONES, H. and MOTT, N. F., 1937, *Proc. Roy. Soc.* A, **162**, 49.
KOPPE, H., 1947, *Z. Naturforsch.*, **2a**, 429.
LANDSBERG, P. T., 1949, *Proc. Phys. Soc.* A, **62**, 806.
PICKARD, G. J. and SIMON, F. E., 1948, *Proc. Phys. Soc.*, **61**, 1.
SEITZ, F., 1940, *Modern Theory of Solids* (New York : McGraw Hill).
SILVIDI, A. A. and DAUNT, J. G., 1950, *Phys. Rev.*, **77**, 125.
SOMMERFELD, A. and BETHE, H., 1933, *Handbuch der Physik* (2nd. ed.) **24**/2, 333.
WIGNER, E., 1934, *Phys. Rev.*, **46**, 1002 ; 1938, *Trans. Farad. Soc.*, **34**, 678.
WOHLFARTH, E. P., 1949, *Phil. Mag.*, **40**, 703.

L. *Note on the Inversion of the Laplace Transform.*

By B. Gross *†.

[Received March 2, 1950.]

When a function $h(\omega)$ is the Fourier transform of a function $H(\sigma)$, then $H(-\omega)$ will be the transform of $h(\sigma)$, the two functions forming a pair. A table of Fourier transforms therefore can be considered as having two entries. The Laplace transform, however, is different, in that the transform is not reciprocal. Thus, if $f(s)$ is the Laplace transform of $F(t)$,

$$f(s)=\int_0^\infty \exp(-st)F(t)\,dt=L\{F(t)\}, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

the knowledge of $f(s)$ will not, in general, enable one to find the function $g(u)$, that is the inverse transform of $F(t)$, and therefore satisfies the relation :

$$F(t)=\int_0^\infty \exp(-ut)g(u)\,du=L\{g(u)\}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

For particular classes of functions, such a formula does however hold. With their aid it is frequently possible to invert the relation (2) and calculate $g(u)$, provided $f(s)$ is known. In spite of its particular nature this formula seems to us to be of considerable importance because it largely extends the scope of existing tables of Laplace transforms *restoring* to this transform up to a certain point, the " pair property " which has been found to be useful for the Fourier transform. The purpose of the present note is to call attention to this formula.

From (1) and (2) it follows that $f(s)$ is the iterated transform of $g(u)$, and therefore (Widder 1946),

$$f(s)=\int_0^\infty \frac{g(u)\,du}{u+s}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This is Stieltjes equation, which is inverted by (Titchmarsh 1937)

$$g(s)=\frac{1}{\pi}\operatorname{Im} f[s\exp(-i\pi)]=-\frac{1}{\pi}\operatorname{Im} f[s\exp(+i\pi)]. \quad . \quad . \quad (4)$$

It has been supposed that all functions are real and s a real variable. The relation (4) obviously resolves the problem because it inverts (2), provided the transform of $F(t)$ is known. The condition of validity is that $g(z)=g[r\exp(i\theta)]$ is analytical for $-\pi<\theta<\pi$, and that

$$\int_0^\infty |g[r\exp(i\theta)]|^2\,dr$$

is bounded for $-\pi<\theta<\pi$ (Gross and Levi 1947).

---

* At present at the Electrical Research Association, on leave of absence from the National Institute of Technology, Rio de Janeiro, Brazil.

† Communicated by Dr. S. Whitehead.

The *result* may be summarized as follows :—

Laplace transform	Original function or inverse transform
$f(s)$	$F(t)$
$F(s)$	$\frac{1}{\pi}\operatorname{Im} f[t\exp(-i\pi)]$

As an *example*, consider the function

$$F(t)=\frac{1}{\sqrt{(\pi t)}}\exp(-k^2/4t),$$

$f(s)$ can be obtained from a table of transforms as

$$f(s)=\frac{1}{\sqrt{s}}\exp -k\sqrt{s}.$$

Since

$$\sqrt{[s\exp(-i\pi)]}=\sqrt{s}\exp(-i\pi/2)=-i\sqrt{s},$$

it follows that

$$f[s\exp(-i\pi)]=\frac{i}{\sqrt{s}}\exp(-ik\sqrt{s})=\frac{i}{\sqrt{s}}(\cos k\sqrt{s}-i\sin k\sqrt{s}),$$

and therefore

$$\frac{1}{\pi}\operatorname{Im} f[s\exp(-i\pi)]=\frac{\cos k\sqrt{s}}{\pi\sqrt{s}}.$$

Therefore

Laplace transform	Original function or inverse transform
$\frac{1}{\sqrt{s}}\exp(-k\sqrt{s})$	$\frac{1}{\sqrt{(\pi t)}}\exp(-k^2/4t)$
$\frac{1}{\sqrt{(\pi s)}}\exp(-k^2/4s)$	$\frac{\cos k\sqrt{t}}{\pi\sqrt{t}}$.

The author is greatly indebted to Dr. S. Whitehead for his helpful advice.

References.

Widder, D. V., 1946, *The Laplace Transform* (Princeton : University Press), 339–341.

Titchmarsh, E. C., 1937, *Introduction to the Theory of Fourier Integrals* (Oxford: Clarendon Press), 317–319.

Gross, B. and Levi, B., 1947, *Matematicae Notae*, **6**, 213.

LI. *The Fluorescence of Silver Halides at Low Temperatures.— Part II. Mixed Crystals of Silver Halides**.

By G. C. FARNELL, Ph.D., P. C. BURTON, B.Sc.,
and R. HALLAMA †‡.

[Received March 13, 1950.]

[Plates XIV.–XVII.]

ABSTRACT.

A description is given of the fluorescence emission spectra of mixed crystals of the silver halides. The results are interpreted from a knowledge of the emission and absorption spectra of the pure halides, and the extents of mixed crystal formation between them.

Some limited results for the temperature-variation of fluorescence efficiency are quoted, which are later shown to be quite compatible with theoretical predictions from a simple model put forward to explain the general result of pure crystal fluorescence.

§ 1. INTRODUCTION.

THE fluorescence spectra of pure silver halides at low temperatures have been described in a previous paper (Part I.). It was felt that an examination of the emission spectra of mixed crystals of the pure silver halides would also provide results of some interest. In particular the silver bromo-iodide system, which, for low iodide content, gave a strong green emission at 77° K., already mentioned briefly in Part I., seemed worthy of further investigation. The majority of results to be described in this Part are concerned with the emission spectra of mixed crystals, but some quantitative results on the variation of fluorescence intensity with temperature of silver bromo-iodide will be described and discussed.

§ 2. MIXED CRYSTAL PREPARATION.

Precipitated mixed crystals were made by pouring a mixed potassium halide solution into $AgNO_3$ solution, so that the differences of solubility of the different silver halides could not greatly affect the formation of mixed crystal. If the $AgNO_3$ solution is poured into the halide solution there is a tendency for the least soluble silver halide to precipitate first. A.R. quality reagents were employed, previous work having shown that the impurities present in them were not responsible for, nor had any influence on, the characteristic emissions of the halides at low temperatures.

---

* Communication No. 1315 H from the Kodak Research Laboratories.

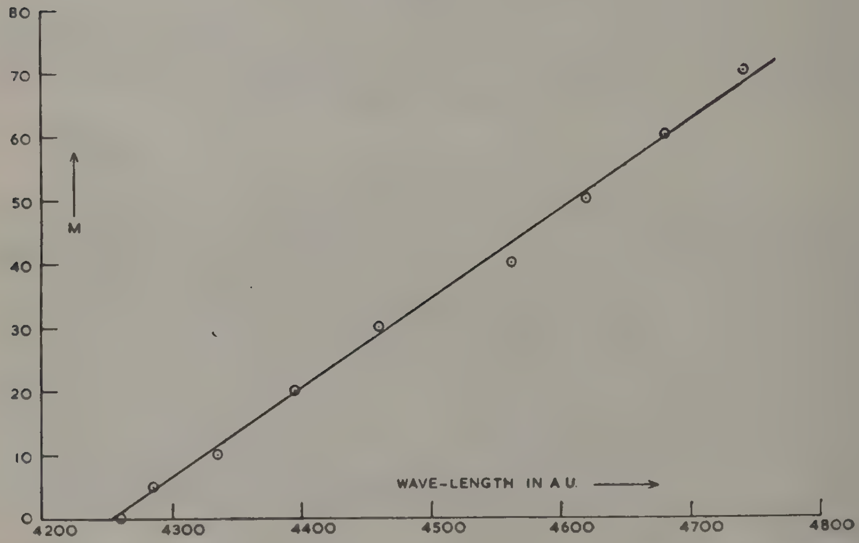
† Now at the University of Helsinki.

‡ Communicated by the Authors.

X-ray powder diffraction patterns were obtained for almost every specimen. It was deduced from these patterns that over ranges in which genuine mixed crystal formation was possible, the proportions of halides in the solutions represented with fair accuracy the actual composition of the mixed crystal. Thus the proportions quoted later are those of the halides in the precipitating halide solution, and may, when only one phase results, be taken as representing the composition of the mixed crystal.

Each of the three possible pairs of silver halides were precipitated together in the proportions of 5, 10, 20, . . . 90, 95 mol. per cent of one component in the other. In addition to these series, mixed crystals were also made of 5, 2, 1, 0·1, 0·01, 0·001 and 0·0001 mol. per cent of AgI in AgBr and AgCl.

Fig. 1.



Shift of short wave-length end of emission band of bromo-chloride series with increase of AgBr content. M=mol. per cent AgBr.

§ 3. Bromo-Chlorides.

The bromo-chlorides are known to form a continuous series of mixed face-centred cubic crystals (Wilsey 1925). This result was confirmed by our own analysis. The emission spectra of the series are illustrated in Pl. XIV. As they were recorded at a specimen temperature of 77° K., only the red "silver" band of AgBr appears at high AgBr content: the characteristic green fluorescence of pure AgBr only just commences at 77° K. The regular shift of the short wave-length edge of the emission band towards longer wave-lengths with increasing AgBr content is illustrated graphically in fig. 1.

It is clear from spectral sensitivity measurements on bromo-chloride emulsions that the absorption edge moves over regularly from the AgCl position to the AgBr position on changing the bromo-chloride composition in that direction. It is therefore reasonable to suppose that the fluorescence emission spectra of a bromo-chloride series should show a regular change from that characteristic of AgCl to that characteristic of AgBr. The recordings illustrated in Pl. XIV. show this fairly clearly, even though they were made at a specimen temperature of 77° K. The characteristic emission of pure AgBr is known to lie in the range 5000 A. to 5500 A. at 77° K., and the spectra visible up to 70 mol. per cent AgBr tend towards this final position. Alternatively one can say that the spectra all appear to represent the emission of AgCl, but the short wavelength edge is progressively shifted to longer wavelengths by the self-absorption of the AgBr constituent (the absorption edge of AgBr lies at a somewhat longer wavelength than that of AgCl). It is important to remember, however, that the allowed energy band structure of the mixed halide is formed by interaction between silver and chloride and bromide ions, and that AgBr does not exist as a separate entity in the mixed crystal. For this reason the former explanation of the change in position of the spectra seems to be preferable.

§ 4. CHLORO-IODIDES.

X-ray powder diagrams showed that at the most there was only mixed crystal formation between AgCl and AgI to the extent of about 5 mol. per cent AgI in AgCl and vice versa. It is in fact well known that chlorides and iodides of identical cation only form small ranges of mixed crystals, owing to the large difference in size of the chloride and iodide ions (Grimm 1925). This lack of mixed crystal formation is evident from the appearance of the emission spectra shown in Pl. XV. Generally the emission spectrum of a particular specimen appears as the photographic sum of the emissions of the separate halides. An apparent exception, however, is the weakening of the AgCl emission for AgI contents of 0·001, and 0·01 and 0·1 mol. per cent AgI, particularly in the range 5400 A. to 5000 A. The spectra were all recorded under constant exposure conditions.

The narrow short wavelength band characteristic of pure AgI became visible at 10 mol. per cent AgI, and, like the main AgI band, showed no change of position with variation in precipitate composition. The X-ray powder diagrams showed that the structure of the AgI component was practically pure cubic (zincblende-type), very little of the hexagonal (wurtzite-type) phase being present.

§ 5. BROMO-IODIDES.

The results obtained with this series were rather more complex than those described in the preceding sections. Previous investigators (Wilsey 1925, Barth and Lunde 1926) have found that mixed crystals are formed at both ends of the range: the AgBr structure (face-centred cubic) accepts up to about 30 mol. per cent AgI, but the AgI structure (zincblende or wurtzite type) accepts only a few mol. per cent of AgBr. In the intermediate range both the AgBr-type and the AgI-type of mixed crystal are formed as separate phases. In our series of precipitated specimens the AgI-type of crystal began to appear at about 20 mol. per cent AgI, and could not be distinguished from pure AgI.

The emission spectra of our series at 77° K. are illustrated in Pl. XVI. Specimens of AgBr and AgBr containing 0·0001 and 0·001 mol. per cent AgI gave identical spectra, viz., the weak red band (5800 A. to 6450 A.) due to the presence of free silver, a phenomenon which has been previously described (Part I). At 0·01 mol. per cent AgI, however, a weak green band appeared (5050 A. to 5700 A.) whose intensity increased rapidly with increasing AgI content up to 4–5 mol. per cent AgI. The spectra illustrated were recorded with different exposure times and are thus deceptive as regards intensities. Beyond 5 mol. per cent AgI the intensity of the green band decreased and at the same time the band moved to slightly longer wavelengths (5130 A. to 6030 A.). The intensity decreased very suddenly beyond 30 mol. per cent AgI, but the band remained faintly visible up to 50 mol. per cent AgI. The main emission band of pure AgI became visible at 20 mol. per cent AgI and steadily increased in intensity up to 100 mol. per cent AgI. Meidinger (1939 b, 1940) has recorded essentially the same position for the intense green band when emitted by bromo-iodide gelatin emulsion layers of fairly low AgI content. He has also confirmed the fact that at around 88° K. maximum fluorescence efficiency is obtained with a specimen containing about 5 mol. per cent AgI. He actually measured the quantum efficiency of fluorescence of a gelatin suspension of AgBr+4 mol. per cent AgI at low temperatures, and obtained a value of about unity.

The emissions of AgBr+4 mol. per cent AgI at 77° K. and 20° K. are compared in Pl. XVII. Meidinger's result for the quantum efficiency was confirmed in so far as there was no difference in fluorescence efficiency at the two temperatures. It is reasonable to suppose that the temperature-dependent fluorescence of a pure solid will rise to a maximum efficiency of unity. It will be seen from Pl. XVII. that there is a very slight shift of the band towards longer wavelengths on reduction of the temperature. There was no observable shift of the band on raising the specimen temperature from 77° K. to 165° K.; however, the fluorescence was quenched at the latter temperature. Meidinger confirms the absence of a shift on raising the temperature from 88° K.

The green band could be excited by radiation of any wavelength up to the short wavelength edge of the emission band, the latter being known to also represent the long wavelength absorption edge of bromo-iodide.

The sudden fall of intensity of the green band beyond 30 mol. per cent AgI is presumably connected with the fact that almost pure AgI is precipitating out as a separate phase, but it is probably due more to a preferential absorption of the exciting radiation by the AgI than to a marked fall in the proportion of the bromide-type phase being formed. In the region of excitation silver iodide possesses a very much higher absorption coefficient than silver bromide. The second factor has probably some influence, however.

The fact that the green emission can be activated by such small percentages of AgI and that its position does not markedly shift over a wide range of AgI contents suggests that the emission is localized in the iodide ion. There is no tendency to approach the position of the emission band of pure AgBr, as in the case of bromo-chloride mixtures. The green bromo-iodide band probably has the same origin as the pure AgI band, but is displaced owing to the different environment of the iodide ion. In both types of the AgI crystal it has four silver ions as nearest neighbours, but in the bromide-type crystal it has six.

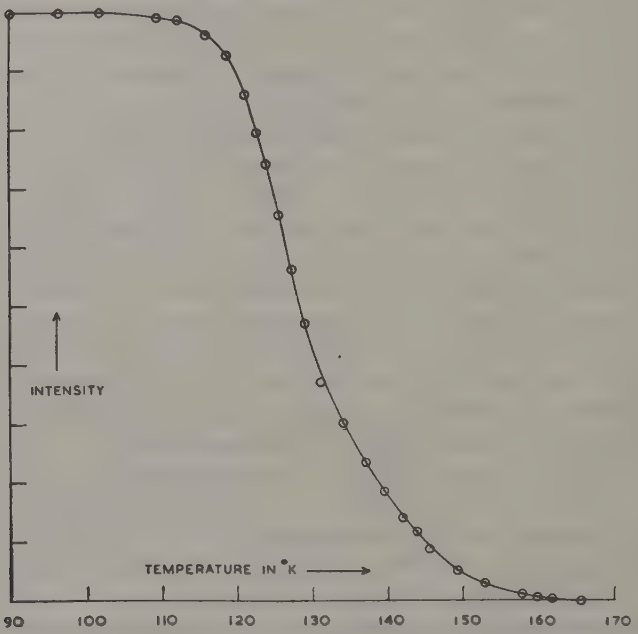
If the emssion is seated in the iodine ions of the mixed crystal it is essential that absorbed quanta should finally give rise to excited iodide ions. Silver iodide has a much higher absorption coefficient throughout the ultra-violet region than silver bromide, and thus a more than proportional amount of incident radiation will be absorbed by iodide ions. The remaining radiation not absorbed by them will be absorbed by bromide ions, and it must be assumed that the positive hole remaining after excitation travels a very short distance to a nearby iodide ion where it becomes trapped. Positive holes are not believed to be mobile over long distances in AgCl and AgBr at low temperatures (Mott and Gurney 1940) but it is quite possible that they can travel the few lattice spacings necessary for them to reach an iodide ion. Evidence in support of this contention of limited positive hole movement at low temperatures is provided by the observations on specimens of $AgCl+MnCl_2$ (described in Part I.), which showed a change in fluorescence colour from blue-green to red on reduction of the exciting intensity. It is impossible to explain the transfer of excitation from chloride ions to manganese ions without assuming a limited positive hole movement.

The fall in intensity of the green band at 77° K. for AgI contents higher than 5 mol. per cent may be due to a decrease in the localization of the iodide ion levels, *i. e.* the concentration of iodide ions reaches a level such that, with the bromide ions, there exists a more uniform energy band structure of the crystal. The probability of the excited electron returning to the ground state without fluorescence, as in AgBr at 77° K. or as a large number of electrons do in AgI, will be increased to a significant value.

§ 6. THE VARIATION OF FLUORESCENCE INTENSITY WITH TEMPERATURE.

The method employed was to focus the emission of a layer of silver halide, mounted on the outer surface of a black painted brass cylinder, on to a photovoltaic cell. The cylinder was allowed to warm up from 77° K. in the open air (protected from draughts) and the temperature of the specimen followed with a copper–constantan thermocouple. The temperature-variation of the cylinder (it required ten minutes to reach room temperature) was very reproducible. From a knowledge of the time which had elapsed since the withdrawal of the cylinder from the liquid nitrogen, the temperature could be predicted to within one or two degrees. The relative fluorescence intensity was measured in terms of a galvanometer deflexion, it having been ascertained that the measured deflexion was always proportional to the fluorescence intensity. The above simple method could not be employed for temperatures below 77° K. owing to the formation of solid air on the specimen.

Fig. 2.

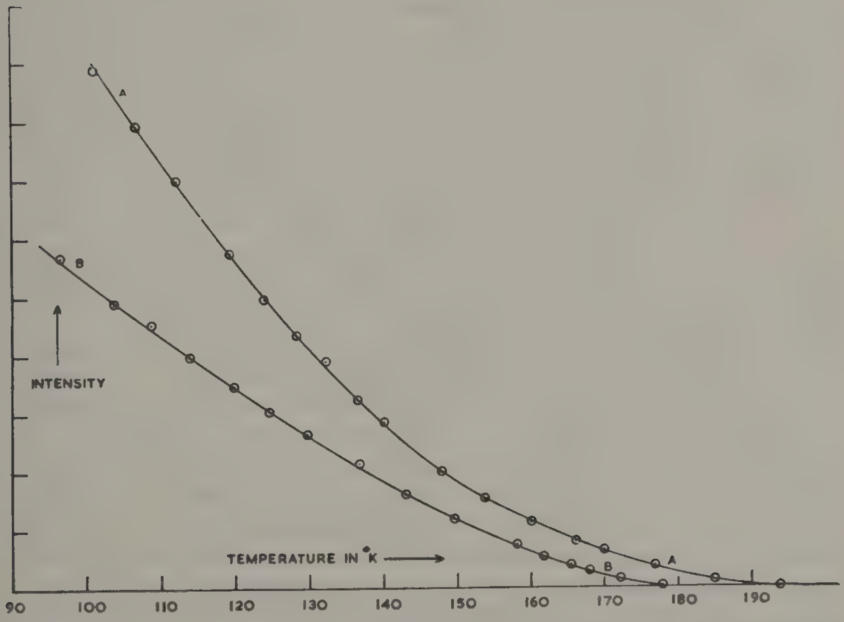


Variation of fluorescence intensity with temperature of AgBr+4 mol. per cent AgI. Intensity in arbitrary units.

As the sensitivity of the photovoltaic cell showed marked variation in the visible spectrum, and the emission band of AgCl shifts with temperature, it was only possible to obtain really useful results for a bromo-iodide specimen, and some limited results for AgI specimens for which the maximum fluorescence efficiency occurs at temperatures below 77° K. An attempt was made to examine the variation of fluorescence intensity of AgCl in narrow spectral regions with the aid of the photomultiplier arrangement described in Part I., but the intensities were too low to provide results of any accuracy.

The variation of fluorescence intensity of AgBr+4 mol. per cent AgI is illustrated in fig. 2. The curve was perfectly reproducible provided that a fresh specimen of precipitate was employed for each determination. The change of specimen was required as the electrons trapped at the low temperature always led to the formation of some "print-out" silver on warming up. This "print-out" silver activated a characteristic weak red emission, and also led to some screening of the silver halide. The curve was perfectly reproducible with the same specimen if the temperature was not allowed to rise above 130° K., which agrees with an observation by Meidinger (Meidinger, 1939a) that no photolytic silver is formed when a bromo-iodide emulsion is exposed at temperatures below 133° K. The flattening out of the curve to a constant high intensity for temperatures below 110° K., and the maintenance of this intensity level down to 20° K. (as already illustrated in Pl. XVII.) indicates that the maximum quantum efficiency of fluorescence obtainable by temperature reduction has been reached.

Fig. 3.



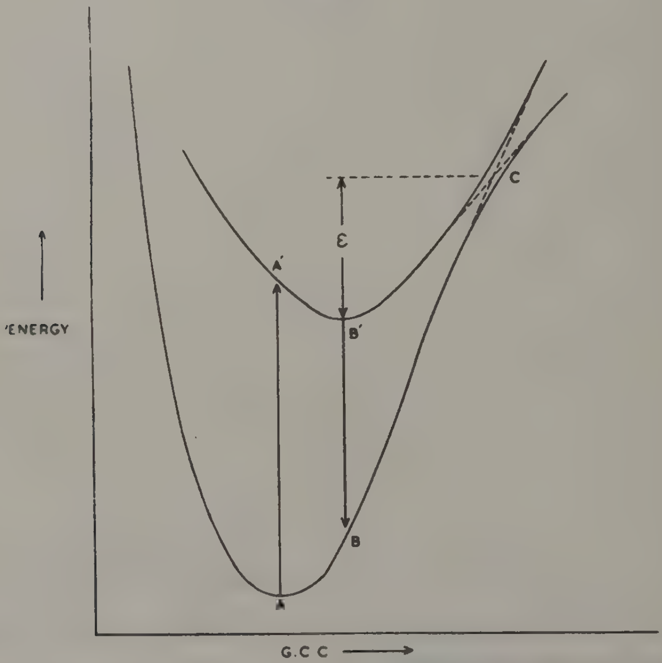
Variation of fluorescence intensity with temperature of two AgI specimens.
Curve A—a hexagonal specimen.
Curve B—a mixed specimen, containing rather more cubic than hexagonal AgI.
Intensities on same arbitrary unit scale for both specimens.

Fig. 3 illustrates the variation of fluorescence intensity with temperature of two AgI specimens, one predominantly hexagonal, the other mainly cubic. The measurements were made under precisely the same geometrical and optical conditions. It will be seen that over the temperature range investigated the hexagonal type possesses a higher fluorescence efficiency.

§ 7. DISCUSSION OF THE RESULTS.

Up till now the emission process has been pictured purely as the recombination of an electron with a positive hole, but in order to discuss the temperature-variation of fluorescence intensity a more detailed picture of the emitting centre is required. We make use of a model which has been employed by several investigators (von Hippel 1936, Mott 1938, Seitz 1939) in considering various problems, and which predicts that a pure solid will fluoresce at a sufficiently low temperature. The energy of the emitting centre depends on the configuration of the central and neighbouring atoms or ions, and although many coordinates are needed to specifiy this configuration, for the purposes of qualitative discussion we shall imagine that only one coordinate is required, as in the case of diatomic molecules. Fig. 4 shows the energy of the normal and excited states of the centre as a function of the configuration

Fig. 4.



A model for the emitting centre in silver halides (after von Hippel, Mott and Seitz). G.C.C. is a generalized configurational coordinate which defines the positions of all ions at once.

coordinate. As the stable configuration of the lattice in the neighbourhood of the centre is different when the latter is in its normal and excited states, the minima of the two curves do not coincide. Moreover, as the excited state approximately corresponds to the transfer of an electron from a negative to a positive ion, the production of a given lattice displacement requires less energy in the region of excitation than it does when the lattice is in its normal state. This is represented by the degree of curvature of the upper curve being less than that of the lower.

Normal optical transitions between the ground state and excited state take place according to the Franck-Condon principle ; the re-arrangement of the ions surrounding the absorbing ion takes place after the electronic transition. The excess energy, corresponding to the difference between A′ and B′, is dissipated in the form of heat, *i. e.* it increases the vibrational energy of the lattice. It is to be noted that the transition indicated in fig. 4 corresponds to absorption in the tail of the band, an exciton, namely an electron bound in the field of a positive hole, being formed. Absorption of shorter wavelengths will raise electrons directly to the conduction band, but, provided that they are not trapped by some special singularity, they will ultimately return to excited states.

It is is advisable to remember at this point, however, that the energy level arrangement shown in fig. 4 is only applicable to a stationary exciton ; the movement of an exciton is too rapid to allow its surrounding ions to achieve equilibrium. A stationary exciton is probably formed in a pure silver halide after excitation by a positive hole becoming trapped at a positive ion vacancy (the latter being definitely known to be present), an electron being subsequently captured in the field of the positive hole. Thus it may not be strictly accurate to speak of the fluorescence of a pure solid, as some defect is required in order that the above mechanism be effective, but rather of a fluorescence which is not due to any specific impurity. All references to pure solid fluorescence in this work should therefore be read with this provision in mind. In the case of the bromo-iodide system positive holes are trapped at the sites of iodine ions to form stationary excitons after electron capture.

The excited system may return to the ground state in two ways, either by making a direct jump with the emission of fluorescence, or transferring to a high vibrational level of the ground state from which it may gradually return with the transfer of energy to the lattice. This latter process requires a little further explanation. In fig. 4 it will be seen that at an energy ϵ above the minimum of the excited state, the ground state and excited state curves closely approach each other. Theoretically, actual crossing is forbidden, but there is an appreciable chance of the excited state changing over, with the emission of one or two vibrational quanta, to a high vibrational level of the ground state. It will be seen immediately that this latter process is temperature-dependent ; if the excited system cannot receive energy ϵ from the lattice, a return to the ground state can only be achieved by emission of fluorescence. Thus at high temperatures there is competition between the dissipation process and the fluorescence process, whereas at low temperatures only the fluorescence process can occur.

The above ideas can be formulated mathematically. If F is the probability per second of the fluorescence process (which will be assumed to be independent of temperature), and D the corresponding quantity for the dissipation process, the quantum efficiency of fluorescence is given by

$$\eta = \mathrm{F}/(\mathrm{F}+\mathrm{D}) = 1/(1+\mathrm{D}/\mathrm{F}), \quad \ldots \ldots \quad (1)$$

D being a function of temperature. D will contain the ordinary Boltzmann factor, $\nu \exp(-\epsilon/k\mathrm{T})$, where ν is the vibrational frequency of the system in its excited state, ϵ has the meaning illustrated in fig. 4, and T is the absolute temperature. A factor β must also be included in D to allow for the fact that the excited state does not necessarily transfer to the ground state every time it reaches the part C; β may possess any value up to unity. Thus (1) may be written

$$\eta = 1/[1+\mathrm{K}\exp(-\epsilon/k\mathrm{T})], \quad \ldots \ldots \quad (2)$$

where $\mathrm{K} = \nu\beta/\mathrm{F}$. For the purpose of comparison of experiment with theory it is to be noted that expression (2) can be transformed to

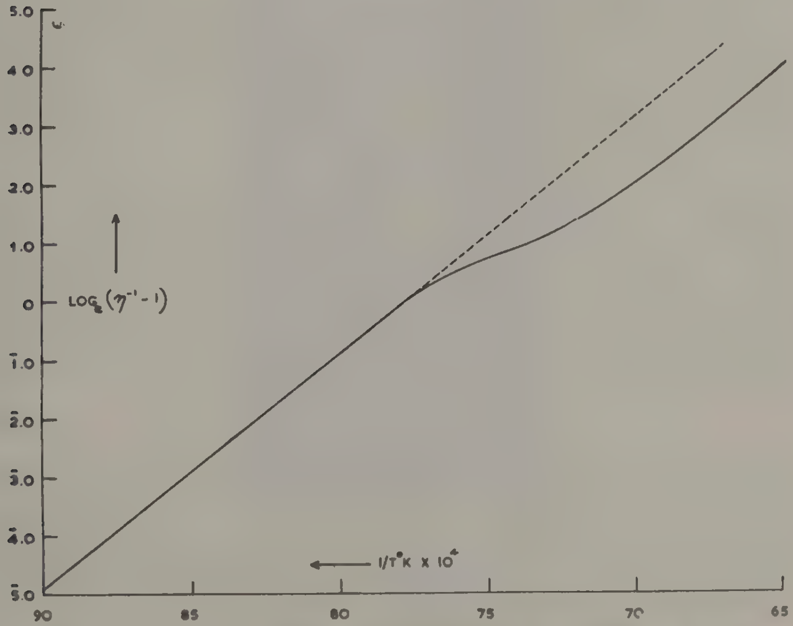
$$\ln(\eta^{-1}-1) = \ln \mathrm{K} - \epsilon/k\mathrm{T}. \quad \ldots \ldots \quad (3)$$

Thus a plot of $\ln(\eta^{-1}-1)$ against 1/T should yield a straight line the slope of which gives ϵ/k, and whose intercept with the $\ln(\eta^{-1}-1)$ axis gives ln K.

If this discussion is correct, the curve of fig. 2 represents essentially a plot of quantum efficiency against temperature; the flat low-temperature part fixes the level $\eta=1$, and we can replot the curve in accordance with expression (3). This is done in fig. 5. It will be seen that for values of η ranging from 1·0 to 0·5 [$\ln(\eta^{-1}-1)$ from $\bar{5}\cdot0$ to 0] and temperatures from 110° K. to 130° K. ($1/\mathrm{T}°\mathrm{K} \times 10^4$ from 90 to 77) a straight line is obtained. Above 130° K., however, the plot corresponds to quantum efficiencies higher than those given by the theoretically predicted straight line. This deviation is to be expected, however, for below 130° K. no appreciable "printing out" occurs, but above this temperature silver commences to be formed. It has been mentioned several times previously that this silver behaves as an activator, and gives rise to a weak red fluorescence. Thus at temperatures above 130° K. the measured intensity is the sum of the green bromo-iodide emission and the red emission. Actually the deviation is of no real consequence as the theory outlined above does not apply when the crystal under consideration is undergoing a photochemical change. So far the experimental results can justifiably be claimed to be in agreement with theory.

The value of ϵ derived from the straight line plot is 0·35 eV., which appears reasonable. The situation with regard to the constant K is less satisfactory. The value derived from the intercept of the straight line plot with the $\ln(\eta^{-1}-1)$ axis is $10^{13\cdot 6}$, which is the experimental value of $\nu\beta/\mathrm{F}$. As no phosphorescence is observed with the bromo-iodides F must be at least 10^5 sec.$^{-1}$; the maximum value of β is unity; and ν will be $\sim 10^{12}$. Thus the maximum theoretical value of $\nu\beta/\mathrm{F}$ is $\sim 10^7$ giving a discrepancy between theoretical and experimental values of $\sim 10^7$.

Fig. 5.



The plot of $\mathrm{l}\eta$ $(\eta^{-1}-1)$, where η=the quantum efficiency of fluorescence, against the reciprocal of the absolute temperature for AgBr+4 mol. per cent AgI. The dashed line represents the course of the plot if expression (2) had been obeyed at higher temperatures.

This discrepancy can be removed, however, by a slight increase in the rigour of the theory. It was assumed earlier that ϵ was independent of temperature, but this can hardly be true considering that the lattice constant and therefore the equilibrium configuration of the ions, is temperature dependent. We shall assume that

$$\epsilon=\epsilon_0+\alpha\mathrm{T}, \qquad (4)$$

where ϵ_0 is the value of ϵ at absolute zero, and α is a small constant of dimensions energy/degree. Referring to expressions (2) and (3) it will be seen that the energy constant derived from the straight line plot is

ϵ_0 (0·35 eV.) and that K has now become equal to $(\nu\beta/F) \exp(-\alpha/k)$. Taking F now equal to 10^8 sec.$^{-1}$, $\beta=1{\cdot}0$, $\nu=10^{12{\cdot}6}$ sec.$^{-1}$, agreement between experiment and theory is achieved if $\alpha=-1{\cdot}8\times10^{-3}$ eV./degree. This is mere speculation, as the values of all other constants appearing in K are uncertain, but the fact that the quantum efficiency-temperature results follow a predicted course, as shown above, does suggest that the essentials of the theory are correct.

It should be mentioned, however, that the same formula for the temperature dependence of efficiency can be obtained on the assumption that ϵ represents the energy needed for the positive hole to escape from the fluorescence centre and travel to a quenching centre. On this basis one could understand why Meidinger, in his investigation of the fluorescence of photographic emulsions, found the quenching temperature of bromo-iodide to vary from one emulsion to another—a very fine-grain emulsion, for instance, showed no fluorescence even at 88° K. We have been unable to confirm this result.

To sum up, we may say that, in general agreement with previous investigators of pure-solid fluorescence (Kudriavzewa 1934, Kröger 1940), we find that the flourescence emission bands of pure and mixed silver halides lie just to the long-wavelength side of their respective absorption edges. Such results as are available on the variation of fluorescence efficiency with temperature are quite compatible with theoretical considerations, and support the simple model put forward to explain them.

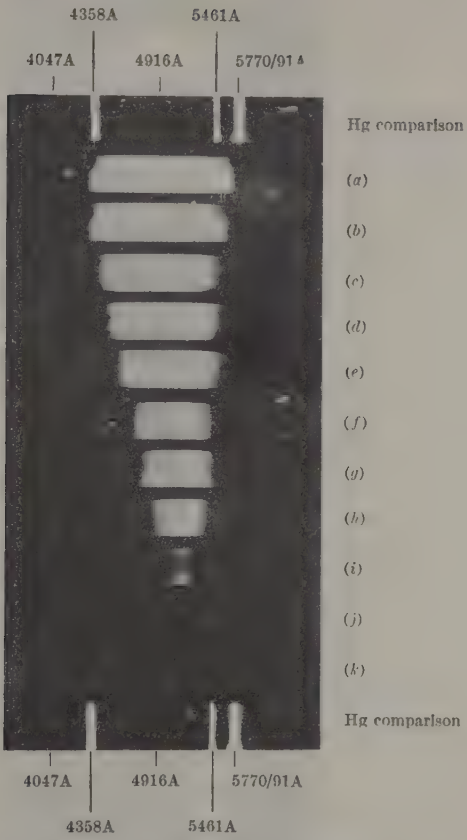
ACKNOWLEDGMENTS.

Our thanks are due to Professor F. E. Simon, C.B.E., F.R.S., of the Clarendon Laboratory, Oxford, for his assistance in enabling us to carry out experiments at liquid hydrogen temperature. We are also indebted to Mr. A. H. Fox, a former member of these laboratories who recorded many of the X-ray powder diagrams of our specimens.

Research Laboratories,
Kodak Limited,
Wealdstone, Harrow, Middx.

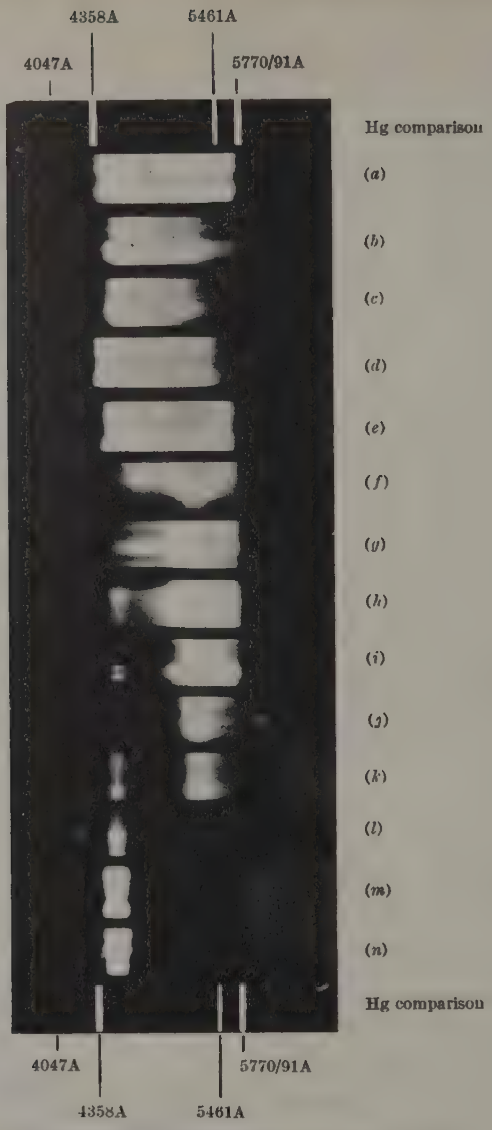
REFERENCES.

BARTH, T. and LUNDE, G., 1926, *Zeit. phys. Chem.*, **122**, 293.
GRIMM, H. G., 1925, *Zeit. phys. Chem.*, **101**, 410.
VON HIPPEL, A., 1936, *Zeit. f. Phys.*, **101**, 680.
KRÖGER, F. A., 1940, *Physica*, **7**, 1.
KUDRIAVZEWA, W., 1934, *Zeit. f. Phys.*, **90**, 489.
MEIDINGER, W., 1939, *Phys. Zeit.*, **40**, 73 (*a*), 517 (*b*); 1940, *Phys. Zeit.*, **41**, 277.
MOTT, N. F., 1938, *Proc. Roy. Soc.* A, **167**, 384.
MOTT, N. F. and GURNEY, R. W., 1940, *Electronic Proccesses in Ionic Crystals* (Oxford: The University Press), p. 138.
SEITZ, F., 1939, *Trans. Faraday Soc.*, **35**, (1), 74.
WILSEY, R. B., 1925, *J. Franklin Inst.*, **200**, 739.



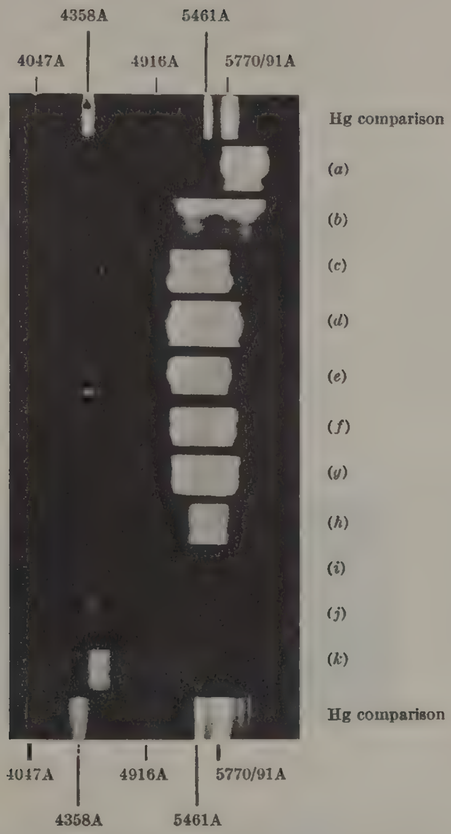
Emission spectra of mixed crystals of AgCl+AgBr at 77° K.

(*a*) Pure AgCl
(*b*) 95 mol. per cent AgCl, 5 mol. per cent AgBr
(*c*) 90 ,, 10 ,,
(*d*) 80 ,, 20 ,,
(*e*) 70 ,, 30 ,,
(*f*) 60 ,, 40 ,,
(*g*) 50 ,, 50 ,,
(*h*) 40 ,, 60 ,,
(*i*) 30 ,, 70 ,,
(*j*) 20 ,, 80 ,,
(*k*) Pure AgBr



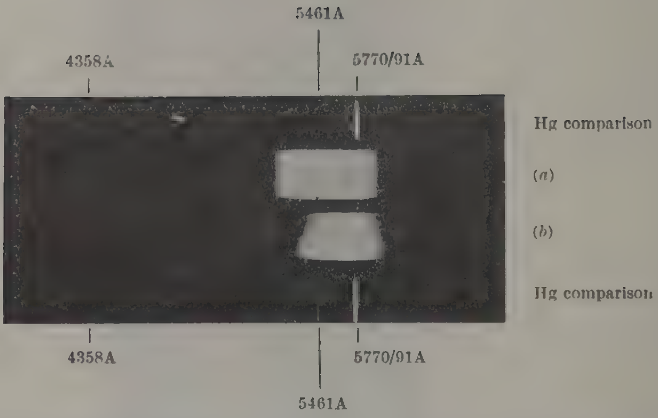
Emission spectra of the chloro-iodide series at 77° K.

(*a*) Pure AgCl
(*b*) 99·999 mol. per cent AgCl, 0·001 mol. per cent AgI
(*c*) 99·99 ,, 0·01 ,,
(*d*) 99·9 ,, 0·1 ,,
(*e*) 99·0 ,, 1·0 ,,
(*f*) 95·0 ,, 5·0 ,,
(*g*) 90·0 ,, 10·0 ,,
(*h*) 80·0 ,, 20·0 ,,
(*i*) 70·0 ,, 30·0 ,,
(*j*) 60·0 ,, 40·0 ,,
(*k*) 50·0 ,, 50·0 ,,
(*l*) 40·0 ,, 60·0 ,,
(*m*) 20·0 ,, 80·0 ,,
(*n*) Pure AgI.



Emission spectra of the bromo-iodide series at 77° K.

(*a*)	Pure AgBr	
(*b*)	99·99 mol. per cent AgBr,	0·01 mol. per cent AgI
(*c*)	99·0 ,,	1·0 ,,
(*d*)	98·0 ,,	2·0 ,,
(*e*)	95·0 ,,	5·0 ,,
(*f*)	90·0 ,,	10·0 ,,
(*g*)	80·0 ,,	20·0 ,,
(*h*)	70·0 ,,	30·0 ,,
(*i*)	40·0 ,,	60·0 ,,
(*j*)	20·0 ,,	80·0 ,,
(*k*)	Pure AgI.	



Comparison of the emission spectra of AgBr+4 mol. per cent AgI at 77° K. (*a*) and 20° K. (*b*). (Recorded with glass spectrograph of higher dispersion.)

LII. *A Determination of the Energy Distribution of Electrons in the Positive Columns of Gas Discharges.*

By C. Haigh and T. Smith*,
Physics Department, Leeds University †.

[Received March 20, 1950.]

Summary.

The electron energies in the positive column of low pressure hot cathode glow discharges have been examined by the magnetic spectrum method.

From measurements of the energy distribution of electrons reaching a positive probe a method of finding the space potential is deduced. With this knowledge a comparison is made of the energy distribution of electrons, in the positive columns of air and argon discharges, determined by the present method and the conventional probe method. Wide discrepancies are observed.

The magnetic spectrum indicates a distribution of the Druyvesteyn form whilst that deduced from probes is Maxwellian. The changing form of the distribution reaching negative probes is proof of the failure at these pressures of the assumptions on which Langmuir probe theory is based. Calculations by Boyd on positive ion sheath formation may give some indication of the mechanisms producing the changing distribution.

§1. Introduction.

The electrical conditions in a discharge are completely specified when the electron and ionic concentrations and energy distributions at every point are known. Of these parameters perhaps the most important is the electron energy distribution since it links many of the macroscopic features of a discharge with the nature of the particles involved. At the same time it is difficult to determine either by experiment or calculation.

The usual way, hitherto employed, of making a direct determination of this energy distribution is the probe method (Langmuir and Mott Smith, 1924), in which the variation of current, with the potential of the probe, is measured.

Langmuir showed that, under ideal conditions, if electrons having a Maxwellian energy distribution reach the probe, a plot of the logarithm of the electron current against the probe potential has a straight line section terminated by a sharp break at the space potential. Unfortunately, many factors, of which the reflection and secondary emission of electrons are the most important, mask the true shape of the characteristic and make the determination, of both the space potential and the energy distribution, difficult (Langmuir, 1924, Loeb 1939).

---

* Now Assistant Lecturer in Physics, Marischal College, Aberdeen.

† Communicated by Professor R. Whiddington, C.B.E., F.R.S.

If a non-linear characteristic is obtained under ideal conditions, it can be analysed by Druyvesteyn's method (1930) in which the energy distribution is obtained from a knowledge of the space potential and the second derivative at every point of the characteristic.

It was to develop a new method of making a direct determination of the electron energy distribution and thereby to test some of the information deduced from probe measurements that the present experiments were carried out. Electrons from the positive column of a discharge have been allowed to pass through a slit in a plane probe, placed parallel to the tube axis, and held at the space potential. The electrons were analysed by the magnetic spectrum method. In this way the energy distribution of the electrons crossing unit area of the discharge per second was obtained. The effect of various types of space charge on the energy distribution can also be investigated, by varying the probe potential, and therefore several assumptions of probe theory can be tested.

§2. Description of Apparatus.

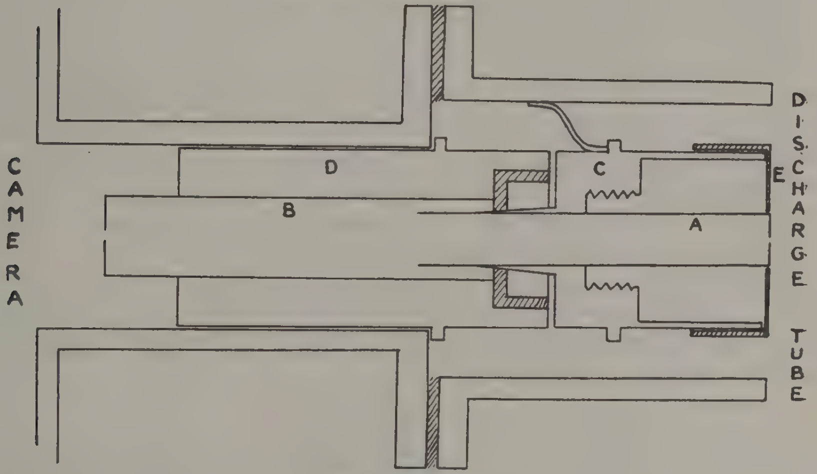
The analysing system is essentially the same as that used in this laboratory for many years (Jones and Whiddington 1928). Its increased size gives a dispersion of 0·4 mm./volt and the different type of slit system produces a photographic trace of a homogeneous beam of electrons having a line width of 0·5 volts.

The slit system (fig. 1) consists of two slits, 0·005 mm. wide pierced in platinum foil at the end of two cylindrical tubes, A, B. The tube nearer to the discharge is 2·43 cm. long and 0·35 cm. in diameter whilst the second tube is 2·70 cm. long and 0·6 cm. diameter. These two tubes fit into two insulated brass cylinders, C, D, so that the slits are 4·7 cm. apart. The electrons are accelerated between the slits before entering the highly evacuated camera.

The system forms a weak electrostatic lens, when potentials are applied between the slits, but for a ratio of final to initial electron energies greater than 6 there is no appreciable aberration produced (Klemperer and Wright 1939). Thus with an accelerating potential of about 200 volts, all electrons having energies less than 30 volts would be brought to the same focus, and the fraction of the initial electrons which pass through the second slit would be independent of their initial velocity. Furthermore, any increase of the accelerating potential would not result in a displacement of the focus and so the fraction would also be independent of accelerating potential. Accordingly, the intensity of the electron beam emerging from the slits was measured whilst the accelerating potential was increased. When the accelerating potential exceeded 200 volts the intensity became constant, showing that the Klemperer and Wright criterion applied.

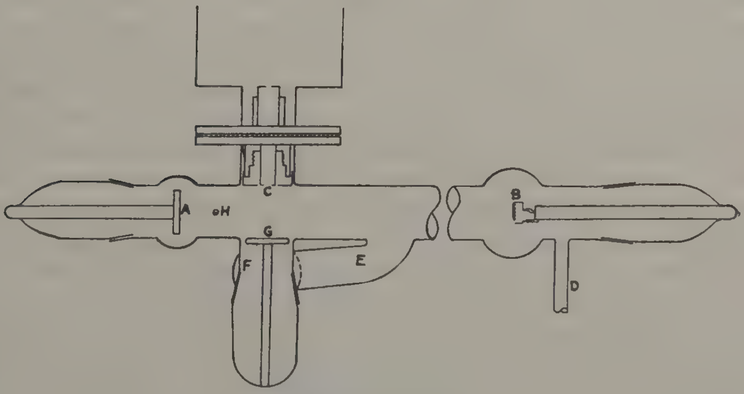
After selection by the slit system the electrons were analysed and recorded on a sensitized film (Brett and Whiddington 1925). The conversion of the photographic blackening, D, to electron intensity, I, was achieved by use of the reciprocity relation ($D=kIt$) over the toe region of the sensitometric curve (Haigh and Smith 1950).

Fig. 1.



The slit system.

Fig. 2.



The discharge tube.

§3. Results.

The discharges examined, so far, have been hot cathode glow discharges at pressures about 10^{-2} mm. Hg. The discharge tube (fig. 2) was constructed throughout of Pyrex tubing of 2 cm. external diameter.

The anode, A, and cathode, B, were fitted to the tube by means of ground joints, the separation between them being 57 cm. The slits, C, were inserted in the side of the tube at a point in the positive column 6·5 cm. from the anode. The gas inlet was through a tube, D, near the filament whilst evacuation was accomplished through a wide tube, E, 16 cm. from the slits. This ensured that the discharge was undisturbed in the region of the slits. The slits could be removed through an opening, F, in the side of the tube. This opening was closed by a glass stopper, G, when the discharge was in progress. A small wire probe, H, was inserted 3·1 cm. from the anode.

The tube potential was maintained by a stabilized power supply giving voltages up to 1000 volts over a range of tube currents of 10 to 100 m.amps. The tube current itself was controlled by variation of the filament heating current. Other potentials were maintained by banks of accumulators and controlled by means of suitable resistances. The circuit allowed variation and measurement of the potentials between the anode and first slit and between the first and second slit.

The ability to reproduce the discharge conditions was demonstrated by the constancy of the characteristic of the small wire probe on different occasions. It was found that careful control of the tube conditions yielded constant characteristics.

Probe Characteristics.

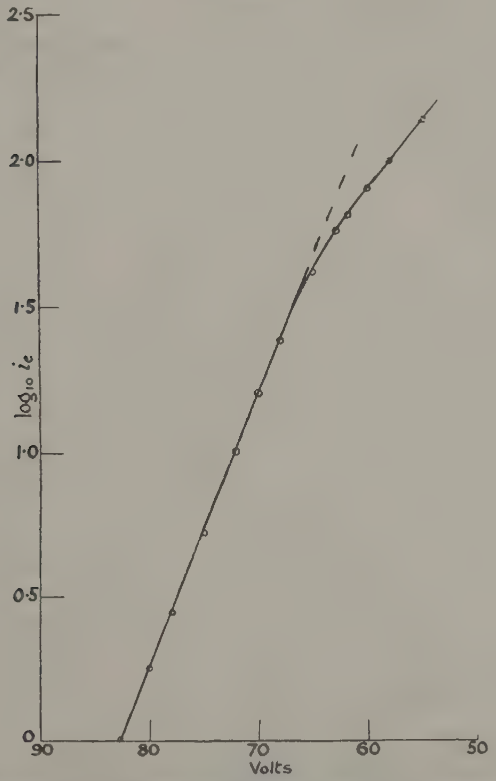
The slit system was of an irregular shape but by placing over the front a cap of copper, or mica, E, (fig. 1) the surface presented to the discharge could be used as a flat probe. The first investigations showed that, although the wire probe characteristic remained constant, those obtained with the slit suffered a gradual variation. If the slits had been cleaned the change in the characteristics was considerable during the first half hour. The change corresponded to a space potential variation of about 4 volts although the actual shape of the semi-log plot remained unchanged. The potential difference between the anode and the space always increased. Thereafter only a small drift in the characteristic was found. This may be readily accounted for by a progressive change in the state of the slit surface. It was necessary to clean the slits regularly and so the discharge was always allowed to run until a constant characteristic was obtained, before taking the energy distribution photographs.

These characteristics when analysed by the conventional methods indicate a Maxwellian energy distribution, as would be expected from the straight portion of the semi-log plot (fig. 3). It is to be noted that the straight line is quite definite, extending over a hundred fold range of probe current. A determination of the space potential from the current reaching a positive probe gives it at a point 6 volts positive to the break point. The determination, however, is rather uncertain (Langmuir and Mott Smith 1924). In the present investigations the space potential may be determined within one or two volts by a method independent of the probe characteristic.

Determination of the Space Potential.

When the first slit is held at a potential positive to the plasma the electron current is limited by a sheath of negative space charge across which there is a rapid change of potential. If the electrons pass through the space charge without collision they will reach the probe with the energy distribution characteristic of the positive column but having gained energy in crossing the sheath.

Fig. 3.

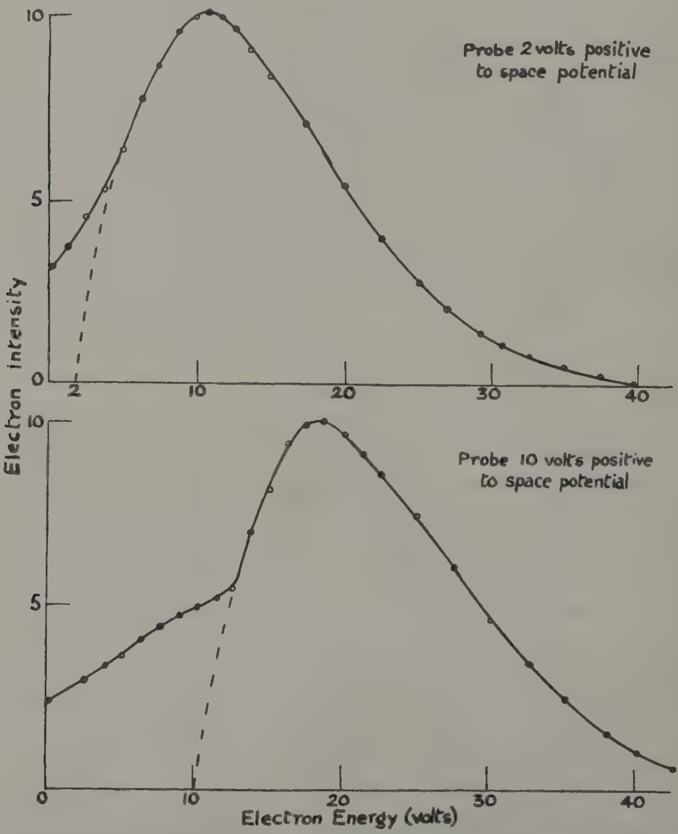


Semi-log plot obtained with a flat probe. Gas, argon ; tube voltage, 600 volts ; tube current, 40 m.amps. ; pressure, 0·025 mm. Hg.

A series of photographs was therefore taken at various potentials relative to the anode. Since the position on the film corresponding to zero energy in the discharge is known accurately, and since the distance on the film can be calibrated in volts, it is possible to measure the space charge interval corresponding to each voltage. A plot of probe potential against space charge interval enables a determination of the space potential to be made. The intensity curves of two photographs from one of the many series which have been obtained are shown in fig. 4. Similar photographs have been obtained with both air and argon discharges.

It can be seen that, when the distributions obtained with the first slit at two different potentials are extrapolated to the abscissa their slowest electrons have an energy corresponding to 2 volts and 10 volts respectively, the potentials of the slit were therefore 2 volts and 10 volts above the space potential, and so the space potential can be found. The low energy electrons which make the extrapolation necessary are made up of secondary electrons emitted from the slit edges, and of electrons which have lost energy by collision in the sheath (Smith, Tait and Whiddington 1949). Little error is introduced by the extrapolation since the form of the distribution in this region is taken from the distribution reaching a negative probe. Here no extrapolation is needed and the uncertainty introduced by the slightly different distribution does not amount to more than 1 volt. This method indicates that the space potential in both air and argon discharges, at these pressures, may be regarded as being about 4 volts positive to the break point.

Fig. 4.



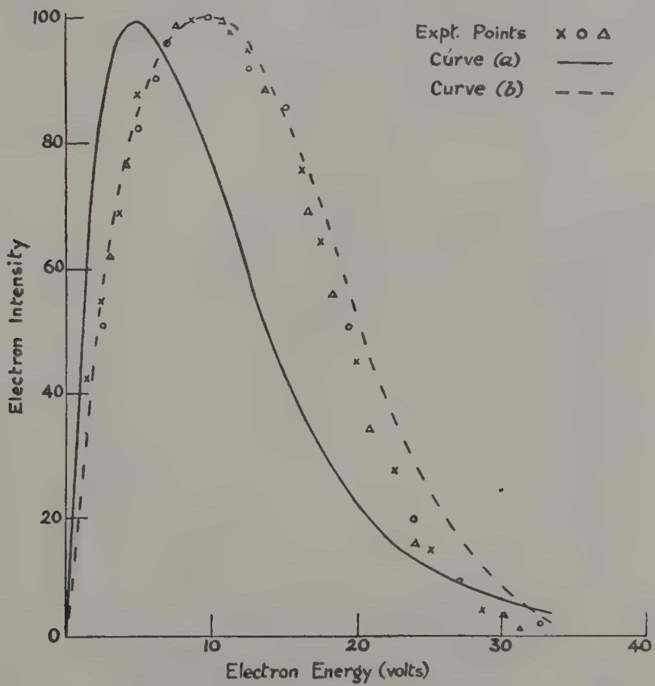
Energy distribution reaching positive probes. Gas, air ; pressure, 0·018 mm.Hg ; tube voltage, 600 volts ; tube current, 40 m.amps.

The Energy Distribution.

Having determined the space potential, the energy distribution photographs were obtained with the first slit held at that voltage, when the electrons striking it have the undisturbed energy distribution of the plasma.

Points from three separate determinations of the energy distribution in an argon discharge (pressure, 0·025 mm. Hg; tube current, 40 m.amps.; tube voltage, 600 volts) are given in fig. 5. It can be seen that there is close agreement between the different determinations.

Fig. 5.



Electron energy distribution in argon (pressure, 0·025 mm. Hg; tube current, 40 m.amps.; tube voltage, 600 volts).

Experimental points; × ○ △

(*a*) Maxwellian flux distribution E=7 5 volts.

(*b*) Druyvesteyn flux distribution E=11·0 volts.

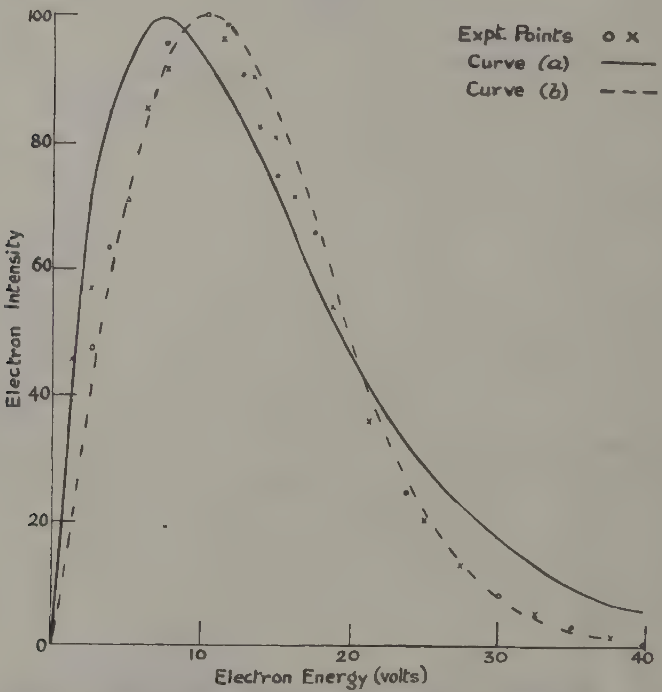
For this discharge, an analysis of the probe characteristic obtained with the first slit, gives the energy distribution of the electrons as Maxwellian with an average energy of 7·5 volts. For a Maxwellian distribution the number of electrons crossing unit area per second having energies in the range V to V+dV is given by

$$N=AV \exp(-1{\cdot}5\, V/E)\, dV. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The flux distribution corresponding to E=7·5 volts is given in fig. 5 (Curve *a*). There is considerable discrepancy between the flux distribution expected from the probe readings, and that actually found, and it is found that no Maxwellian distribution can be fitted to the experimental curve.

The motion of an electron in an electric field is not governed by the simple laws from which the Maxwellian distribution of energies is deduced. The existence of free paths varying with energy and energy losses which are a function of the electron energy have to be considered. Owing to the continual gain and loss of energy in the field, electrons are not in thermal equilibrium with the gas molecules or atoms. Thus in discharge plasma the energy distribution cannot be of the Maxwellian form unless other mechanisms produce further energy exchange. The most important mechanism of this type is electron interaction, but for the electron concentrations obtained in this type of discharge its effect is negligible.

Fig. 6.



Electron energy distribution in air, (pressure, 0·018 mm. Hg; tube current, 40 m.amps.: tube voltage, 600 volts).

Experimental points; ○ ×

(*a*) Maxwellian flux distribution E=10·8 volts.

(*b*) Druyvesteyn flux distribution E=11·0 volts.

The full derivation of the distribution function has only been completed for the case of He (Smit 1936) but a simplified solution has been given by Dryvesteyn (1930).

The flux distribution corresponding to a Druyvesteyn energy distribution of average energy E is given by

$$N=AV \exp(-0{\cdot}55\, V^2/E^2)\, dV \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and it was found that a curve of this form could be fitted approximately to the photometer curve if $E=11{\cdot}0$ volts (fig. 5). Even so it is evident that there is a dearth of electrons at energies higher than the mean.

Similar results have been obtained with air discharges. For a discharge (pressure, 0·018 mm. Hg ; tube current, 40 m.amps. ; tube voltage, 600 volts) the measured flux distribution is given in fig. 6 together with that deduced from probe measurements. Again a considerable difference is obtained between the two distributions and again the distribution obtained by the present method can be represented approximately by a Druyvesteyn flux disuribution.

4. Discussion of Results.

Linear semi-log plots have been obtained frequently from probe measurements under widely differing discharge conditions where the existence of a Maxwellian energy distribution has been doubtful. The factors leading to error have been examined (Loeb 1939), but particular attention may be drawn to the effect of the reflection and secondary emission of electrons from the probe surface. It is interesting to note that in argon discharges the shape of the characteristic in the region of large negative potentials was seriously affected by these factors (Uyterhoeven 1929) yet the electron energy distribution deduced from the semi-log plot was still Maxwellian (Uyterhoeven and Harrington 1930). In the present method errors due to these causes are eliminated, for, although electrons are reflected and secondary electrons are produced, they are only recorded when the slit is at a potential positive to the plasma. Similarly, because in the present method the energy distribution is determined when the slit is at the space potential all the factors which introduce error into probe determinations, due to their effect on the space charge, are obviated.

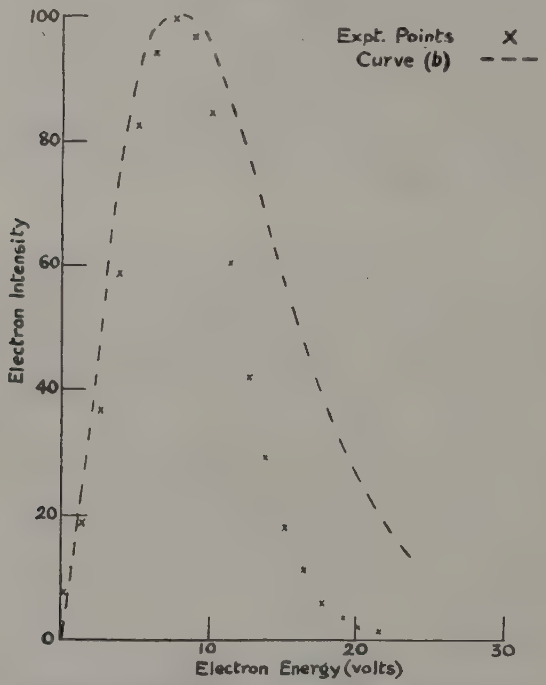
Confirmation of the form of the energy distribution obtained is given by recent calculations of Deas and Emeleus (1949) from the values of the Townsend coefficient, α, with field strength, X, at various pressures, p. Their results show that in the case of argon discharges the experimental variation of α/p with X/p necessitates the assumption of an energy distribution function with a paucity of electrons having energies much in excess of the mean.

Finally, some subsidiary observations bearing on the validity of the probe method of finding energy distributions may be noted. In the derivation of the theory it is assumed that no exchange of electron energy occurs within the sheath. Under these conditions the number of electrons N in an energy range V to V+*d*V, reaching a probe against a negative potential V_0 is given by

$$N = A(V - V_0) f(V) dV. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Where the volume distribution function outside the space charge sheath is $f(V)$ (Eakin and McCrea 1940).

Fig. 7.



Energy distribution reaching negative probe.

(*a*) Probe 8 volts negative to space potential.

(*b*) Distribution according to equation (3).

Accordingly the energy distribution was obtained with the first slit held at voltages negative to the space potential. The distribution obtained in the case of the argon discharge, with the slits held 8 volts negative to the space potential, is shown in fig. 7 together with the distribution calculated from equation (3), when $f(V)$ is assumed to be Druyvesteyn of average energy 11·0 volts.

The increased discrepancy between the experimental and calculated distributions indicates the inadequacy under these conditions of one of the assumptions of probe theory.

Recently Boyd (1950) has used a small screened probe in order to separate the components of the probe current and determine the energy distribution of both the electrons and positive ions. He claims that with the conventional probe, if the positive ion current does not become saturated, the field from the probe modifies the energy distribution and concentration in the probe neighbourhood until, if the probe voltage is sufficient, a positive ion sheath forms. He obtains an electron energy distribution with few electrons having more than 15 volts energy.

ACKNOWLEDGMENTS.

We wish to acknowledge our debt to Professor Whiddington for his continual interest and advice during the course of the investigations and also we would like to take this opportunity of expressing our thanks to all those who have helped in any way. Financial assistance has been afforded to both authors by the D.S.I.R.

REFERENCES

BOYD, 1950, *Nature*, **165**, 228.
BRETT and WHIDDINGTON, 1925, *Proc. Leeds Phil. Lit. Soc.*, **1**, 1.
DEAS and EMELEUS, 1949, *Phil. Mag.*, **40**, 461.
DRUYVESTEYN, 1930, *Physica*, **10**, 69.
EAKIN and MCCREA, 1940, *Proc. Roy. Irish Acad.*, **46** A, 91.
HAIGH and SMITH, 1950, *Proc. Leeds Phil. Lit. Soc.*, in the press.
JONES and WHIDDINGTON, 1923, *Phil. Mag.*, **6**, 889.
KLEMPERER and WRIGHT, 1939, *Proc. Phys. Soc.*, **51**, 296.
LANGMUIR and MOTT SMITH ,1924, *Gen. Elec. Rev.*, **27**, 762.
LOEB, 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York).
SMIT, 1936, *Physica*, **3**, 543.
SMITH, TAIT and WHIDDINGTON, 1949, *Proc. Leeds Phil. Lit. Soc.*, **5**, Part III., 207.
UYTERHOEVEN, 1929, *Proc. Nat. Acad. Sci.*, **15**, 32.
UYTERHOEVEN and HARRINGTON, 1930, *Phys. Rev.*, **36**, 709.
WHIDDINGTON and TAYLOR, 1932, *Proc. Roy. Soc.* A, **136**, 651.

LIII. *The Cohesive Energy of Metallic Magnesium.*

By S. Raimes, B.Sc., A.R.C.S., D.I.C.,
Department of Mathematics, Imperial College, London*.

[Received February 7, 1950.]

Summary.

The method of Wigner and Seitz is applied to metallic magnesium, using an ion-core field of the Hartree type. Two alternative self-consistent fields in the metal are considered, and the wave functions and energies are found not to differ significantly in the two cases. The total energy of the metal is obtained as the sum of four terms: the eigenvalue of the Schrödinger equation for the lowest state, and the Fermi, exchange, and correlation energies. The free-electron approximation is used to evaluate the latter two terms, but, in the case of the Fermi energy, both the free-electron value and a modification of this using the effective electronic mass at the zone centre are considered: the effective mass being calculated by a method due to Bardeen. The energy of the valence electrons in a free atom is obtained by a self-consistent field calculation, and, in addition, an estimate is made of the correlation energy of these electrons. It is pointed out that the method employed in the latter case appears to be inapplicable to beryllium, so that the true correlation energy may differ significantly from the estimated value. However, the calculated values of the lattice constant, cohesive energy, and compressibility are found to be in fair agreement with experiment, and a discussion of the effects of the various approximations is given.

§1. Introduction.

The cellular method introduced by Wigner and Seitz (1933, 1934) has been used with considerable success in the calculation of the cohesive energies of the monovalent metals. So far, however, little has been done in this direction for metals of higher valency. Apart from certain work of an essentially approximate nature, the only detailed investigation of the cohesion of a polyvalent metal has been that of Herring and Hill (1940) in the case of beryllium. The present paper describes an attempt to calculate, by means of the cellular method, the lattice constant, compressibility, and cohesive energy of the divalent metal magnesium. The procedure adopted is not the same as that of Herring and Hill, but is a more direct extension of that used by Wigner and Seitz for sodium. The effective electronic mass at the centre of the Brillouin zone is, however, determined by a method due to Bardeen (1938).

---

* Communicated by Professor H. Jones.

There have been two previous attempts to calculate the cohesive energy of magnesium, among other divalent metals; namely, that of Fröhlich (1936), and that of Hellmann and Kassatotschkin (1936). Both treatments involve the cellular method to some extent, but also include additional approximations; and in neither case are wave functions obtained such as those given in this paper. Fröhlich makes use of his approximate analytical method (1936, 1937) of calculating the energies of the alkali metals; an empirical screening factor being introduced into the potential field to account in some degree for the presence of a second valence electron. Although this method applies extremely well to the alkali metals, its extension to the divalent metals must be regarded only as a rough approximation. It is to be noted, however, that the calculation of the Fermi energy, in the latter case, yields a value considerably less than that for perfectly free electrons: this will be further discussed in §5.

Hellmann and Kassatotschkin use what they term the "combined approximation method." The basis of this lies in the use of an artificial ion-core field so constructed that the empirical energies of the valence electrons in a free atom are obtained when the ground state of a valence electron is taken to be a $1s$ state, and the first and second excited states are taken to be $2p$ and $2s$ states respectively. There is a certain arbitrariness in the construction of this field, since two parameters are fitted to three atomic term values, and so, apart from the ground-term, which is made to agree exactly, the correspondence between the theoretical and experimental energies of the atom is not exact: it is postulated, however, that the resulting error in the energy of the metal will be small. The whole reason for this procedure is that the wave functions of the valence electrons in the metal now have no spherical nodes in the neighbourhood of each nucleus, so that an approximation to the energy may be obtained by assuming that they are just plane waves. The results are in fair agreement with experiment, but the effects of the various approximations are rather difficult to assess.

The present work is essentially an investigation of the direct applicability to divalent metals of the method of Wigner and Seitz, and the ion-core field used is of the Hartree type.

§2. THE WAVE FUNCTIONS.

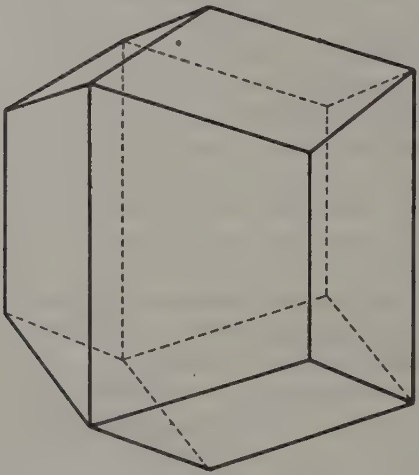
Metallic magnesium has a close-packed hexagonal structure, and a polyhedron drawn around any atom of the lattice, so that the faces bisect perpendicularly the lines joining this atom to its nearest and next nearest neighbours, has the form shown* in fig. 1. This atomic cell has not the high degree of symmetry possessed by those of the cubic lattices, but, for the purpose of calculating the wave function of the lowest electronic state, it is still a reasonable approximation to replace the cell by a sphere of equal volume.

---

* *Cf.* Seitz (1940) p. 331, fig. 4: this figure is believed to be incorrect.

W. J. Yost (1940) has calculated the wave functions of the K and L shells in a free, doubly-ionized, magnesium atom according to the approximations of both Hartree and Fock. The spherically symmetrical ion-core field used in the present work was calculated from the Fock functions, on the assumption that the ion-core remains unchanged in the solid state. Exchange interaction between the valence and core electrons will be neglected both in the metal and in the free neutral atom, so that the error should be small as far as the cohesive energy is concerned.

Fig. 1.



Atomic polyhedron for a close-packed hexagonal lattice.

It will be shown that the wave function $\psi_0(r)$ of the lowest state of a valence electron in the metal is approximately constant throughout most of the volume : hence, as in the case of the monovalent metals, we shall assume that the wave function of an excited electronic state is

$$\psi_j(\mathbf{r})=\psi_0(\mathbf{r}) \exp i\mathbf{k}_j \cdot \mathbf{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\mathbf{k}_j$ is the wave vector. If there are N atoms in the crystal, therefore, the charge density at $\mathbf{r}$ due to electrons with one type of spin only is *

$$\sum_j |\psi_j(\mathbf{r})|^2=\mathrm{N}|\psi_0(\mathbf{r})|^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

* In this paper the following units are used :

$$\text{Unit of energy (Rydberg unit)}=\frac{2\pi^2me^4}{h^2},$$

$$\text{Unit of length (Bohr unit)}=\frac{h^2}{4\pi^2me^2}.$$

In the case of divalent metals there are two alternative self-consistent fields which are plausible on physical grounds, and we shall refer to these as field A and field B :

Field A. It may easily be shown, by regarding the electrons as approximately free, that those with parallel spins rarely approach one another within a distance smaller than r_s, the radius of an atomic sphere. It is therefore reasonable to assume that the presence of an electron with, say, positive spin effectively excludes from an atomic sphere all other electrons with positive spin. Hence we may take the potential field $V_A(r)$ in which this electron moves as the sum of the ion-core field $v_c(r)$ and that due to the charge distribution, within the sphere, of all electrons with negative spin ; *i. e.* using (2),

$$V_A(r)=v_c(r)+2N\int_\sigma \frac{|\psi_0(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}d\tau', \quad . \quad . \quad . \quad . \quad . \quad (3)$$

the integration being over the atomic sphere σ. (The factor 2 appears owing to our choice of units.) This field is similar to that used by Manning and Krutter (1937) in their work on calcium, except that these authors use, instead of $\psi_0(\mathbf{r})$, that portion of an atomic wave function, suitably normalized, which would be enclosed by an atomic sphere. Owing to the form of the wave functions ψ_j, as given in equation (1), the field A becomes that of the valence electrons in the Hartree calculation for the free atom when the atomic sphere becomes infinitely large.

Field B. If correlations between the electronic positions are ignored altogether the potential field $V_B(r)$ within an atomic sphere is the sum of the ion-core field and that due to the charge distribution, in the sphere, of all the electrons : *i. e.*

$$V_B(r)=v_c(r)+4N\int_\sigma \frac{|\psi_0(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}d\tau'. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This is the ordinary Hartree field of a valence electron in the metal.

The above definitions have been given, for convenience in the following discussion, with respect to a single atomic cell, or sphere, but the extension to the crystal as a whole is obvious, since the cells are assumed to be electrostatically neutral and the situation within each cell is the same.

It may be remarked that neither of these fields is correct ; they are both approximations to the Fock field in the metal, but it was decided that field A gives a more plausible representation of the physical picture, and this field was used in the bulk of the work.

Wave functions ψ_0 corresponding to the atomic radii 3·28 and 4·56 Bohr units are given in Table I., as calculated with the two alternative fields. These particular radii were chosen because they include between them the most important part of the range, as can be seen from fig. 3. It is at once apparent that the corresponding functions are very similar, so that we might expect the resulting energy of the crystal to be about the same in the two cases ; and, in fact, this is borne out by calculation.

It will be seen later that the total energy per electron is the sum of the eigenvalue ϵ_0 of the Schrödinger equation for the lowest state and certain other terms. Now, taking the case of $r_s=3{\cdot}28$ only, we find the eigenvalue for field A to be $-{\cdot}66$, and that for field B to be ·05 Rydberg units. However, in the second case, the total energy contains an extra term representing the Coulomb interaction of the electrons, and this is found to have the value $-{\cdot}72$ Rydberg units. The other energy terms remain the same in the two cases, so the net result is that the total energy for field B is ·01 Rydberg units lower than that for field A. This difference, it will appear, is certainly less than the errors expected from other sources.

TABLE I.

r (Bohr units)	ψ_0 for $r_s=3{\cdot}28$		ψ_0 for $r_s=4{\cdot}56$	
	Field A	Field B	Field A	Field B
0·00	5·316	5·236	3·748	3·444
0·08	1·826	1·800	1·287	1·183
0·16	0·270	0·267	0·189	0·174
0·24	−0·401	−0·394	−0·284	−0·261
0·40	−0·690	−0·681	−0·485	−0·447
0·56	−0·548	−0·545	−0·383	−0·354
0·72	−0·348	−0·350	−0·240	−0·224
0·88	−0·173	−0·179	−0·114	−0·109
1·04	−0·036	−0·046	−0·017	−0·021
1·36	0·142	0·131	0·110	0·096
1·68	0·237	0·225	0·173	0·155
2·00	0·282	0·275	0·200	0·183
2·32	0·302	0·298	0·206	0·912
2·64	0·307	0·309	0·202	0·193
2·96	0·308	0·313	0·193	0·189
3·28	0·307	0·313	0·184	0·183
3·60			0·175	0·178
3·92			0·168	0·175
4·24			0·164	0·172
4·56			0·163	0·172

Although the calculations for field B have not been carried to the same degree of self-consistency as those for field A, we may conclude that the final results are not significantly different in the two cases. A similar conclusion was reached by Wigner and Seitz in the case of sodium, using a perturbation method.

We shall confine all further discussion to field A, the procedure for field B differing only in detail.

That portion of ψ_0 lying within a given atomic sphere is obtained by integrating the Schrödinger equation

$$\frac{d^2\psi_0}{dr^2}+\frac{2}{r}\frac{d\psi_0}{dr}+[\epsilon_0-V_A(r)]\psi_0=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

subject to the boundary condition

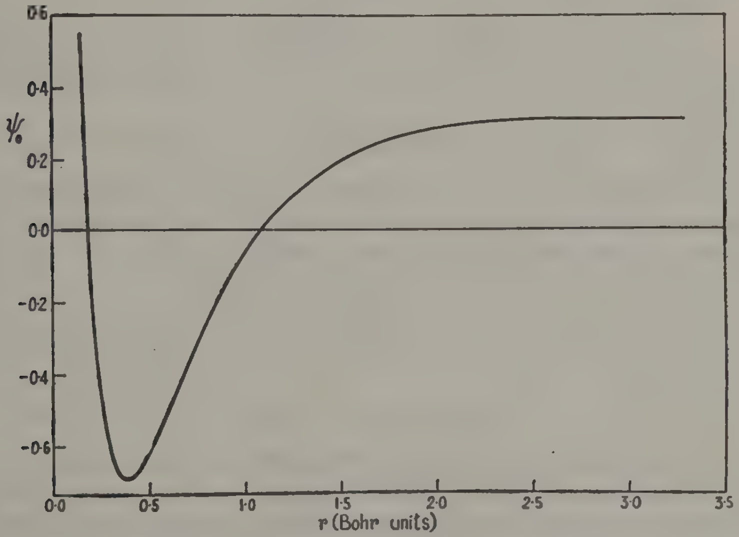
$$\left(\frac{d\psi_0}{dr}\right)_{r=r_s}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

With the form of $V_A(r)$ given by (3), ψ_0 must satisfy the normalizing condition

$$4\pi N\int_0^{r_s} r^2 \,|\,\psi_0(r)\,|\,^2 dr=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In addition ψ_0 must possess two spherical nodes within the atomic sphere.

Fig. 2.



Wave function of the lowest electronic state, for atomic radius 3·28 Bohr units.

Since V_A depends upon ψ_0, a method of successive approximation had to be used in solving equation (5). Furthermore, ϵ_0 had to be chosen so as to satisfy condition (6), with the value of r_s fixed beforehand, and this could be done only by trial and error. Consequently the work involved was much more laborious than in the corresponding calculation for sodium. In the latter case a single integration provided two points on the ϵ_0-versus-r_s curve, while here several integrations, and several recalculations of V_A, were necessary to obtain a single point on the curve. For this reason the solution for a given atomic radius was terminated after two or three recalculations of V_A. This provided a value of ϵ_0 to three significant figures, the third being, however, rather unreliable. Solutions of this accuracy were obtained at six points covering the most important part of the range of r_s, and the resulting curve is shown in fig. 3.

The general form of the wave functions ψ_0 is shown in fig. 2, which represents that for $r_s=3{\cdot}28$ Bohr units, the observed value being 3·35 Bohr units. It is apparent that our previous assumption of the approximate constancy of ψ_0 is justified. The number of intervals actually used in the calculations was about four times the number shown in Table I. All the wave functions given are normalized according to the equation

$$\int_0^{r_s} r^2\psi_0^2 dr=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Although we have been concerned with a single atomic sphere, ψ_0 must, of course, extend periodically throughout the crystal. The Schrödinger equation for an electronic state ψ_j, referred to the metal as a whole, is

$$\nabla^2\psi_j+[\epsilon_j-\mathrm{V}_{\mathrm{A}}(\mathbf{r})]\psi_j=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where $\mathrm{V}_{\mathrm{A}}(\mathbf{r})$ reduces to (3) within any cell. Substituting (1) in (9), we obtain

$$\nabla^2\psi_0+2i\mathbf{k}_j \,.\, \mathbf{grad}\,\psi_0+(\epsilon_j-\mathbf{k}_j^2-\mathrm{V}_{\mathrm{A}})\psi_0=0. \quad . \quad . \quad . \quad . \quad (10)$$

Owing to the approximate constancy of ψ_0 the second term may be neglected to a first approximation, and, within an atomic sphere, equation (10) reduces to (5). Then

$$\epsilon_j=\epsilon_0+\mathbf{k}_j^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

This will be used in the following section, but a fuller discussion will be given in §4.

§3. The Total Energy.

(i) *The solid metal.*

In order to construct a total Hamiltonian for the metal the following simplifying assumptions, which are implicit in the previous work, will be made: (*a*) The atomic nucleii are at rest. (*b*) The inner closed shells of electrons remain the same in the metal as they are in the free atom, and their effect may be described by a spherically symmetrical potential term. (*c*) The ion-cores do not overlap appreciably, and Van der Waals terms may be neglected.

The divalent metallic system then consists essentially of 2N valence electrons moving in a lattice of N doubly-charged positive ions, and the corresponding Hamiltonian operator is

$$\mathrm{H}=-\sum_{i=1}^{2\mathrm{N}}\nabla_i^2+\sum_{i=1}^{2\mathrm{N}}\sum_{a=1}^{\mathrm{N}}\mathrm{U}_a(\mathbf{r}_i)+\sum_{i\neq j}^{2\mathrm{N}}\sum\frac{1}{\mathbf{r}_{ij}}+\sum_{a\neq b}^{\mathrm{N}}\sum\frac{4}{\mathbf{r}_{ab}} \quad . \quad . \quad . \quad (12)$$

Here $-\nabla_i^2$ is the kinetic energy operator of the ith electron; $\mathrm{U}_a(\mathbf{r}_i)$ is the potential energy of the ith electron due to the ath ion; $2/r_{ij}$ is the Coulomb interaction energy between the ith and jth electrons, where r_{ij} is the distance between them; and, similarly, $8/r_{ab}$ is the Coulomb interaction energy between the ath and bth ions.

In the one-electron approximation, when there are N electrons with each type of spin, we may take the total wave function to be of the form :

$$\Psi=\frac{1}{\mathrm{N}!}\begin{vmatrix}\psi_1(\mathbf{r}_1) & \dots & \psi_{\mathrm{N}}(\mathbf{r}_1)\\ \vdots & & \vdots \\ \psi_1(\mathbf{r}_{\mathrm{N}}) & \dots & \psi_{\mathrm{N}}(\mathbf{r}_{\mathrm{N}})\end{vmatrix}\cdot\begin{vmatrix}\psi_1(\mathbf{s}_1) & \dots & \psi_{\mathrm{N}}(\mathbf{s}_1)\\ \vdots & & \vdots \\ \psi_1(\mathbf{s}_{\mathrm{N}}) & \dots & \psi_{\mathrm{N}}(\mathbf{s}_{\mathrm{N}})\end{vmatrix} \quad . \quad . \quad . \quad (13)$$

where the $\mathbf{r}_i$ and $\mathbf{s}_i$ are the position vectors of electrons with positive and negative spin respectively, and

$$\int_{(v)}\psi_i{}^*\psi_j\,d\tau=\delta_{ij},$$

the integration being over the volume v of the crystal.

The total energy of the crystal in this approximation is

$$\mathrm{E}=\int\dots(2\mathrm{N})\dots\int_{(v)}\Psi^*\mathrm{H}\Psi\,d\tau(\mathbf{r}_1,\dots\mathbf{r}_{\mathrm{N}}\,;\,\mathbf{s}_1,\dots\mathbf{s}_{\mathrm{N}}). \quad . \quad . \quad (14)$$

In evaluating this expression use is made of equation (9) and the atomic sphere approximation ; the latter in so far as the charge distribution within an atomic cell is assumed to be spherically symmetrical, so that, as far as points outside are concerned, the charge can be regarded as concentrated at the centre of the cell. Finally, after some calculation, (14) reduces to

$$\mathrm{E}=2\sum_{i=1}^{\mathrm{N}}\epsilon_i-2\sum_{i,\,j}^{\mathrm{N}}\iint_{(v)}\frac{1}{r_{12}}\psi_i{}^*(\mathbf{r}_1)\psi_j{}^*(\mathbf{r}_2)\psi_i(\mathbf{r}_2)\psi_j(\mathbf{r}_1)\,d\tau(\mathbf{r}_1,\mathbf{r}_2). \qquad (15)$$

It is noteworthy that this contains no specific term for the Coulomb interaction energy, the second term being the exchange energy. This is due entirely to our use of field A ; when field B is used an additional Coulomb term does appear, as mentioned in §2.

We may express (15) in the form

$$\mathrm{E}=2\mathrm{N}(\epsilon_0+\epsilon_{\mathrm{F}}+\epsilon_{\mathrm{X}}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where ϵ_{F} is the Fermi energy and ϵ_{X} the exchange energy, each referred to a single electron. If we assume that the ψ_j are approximately plane waves ϵ_{F} and ϵ_{X} can be evaluated by well-known methods. In the case of ϵ_{F} use is made of equation (11). We find

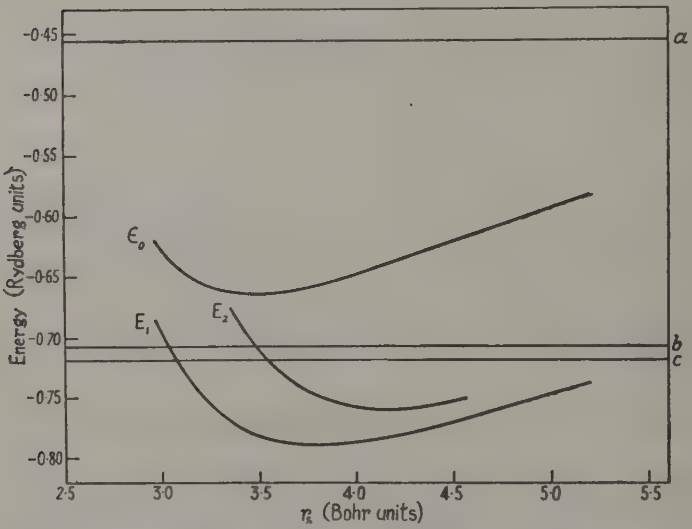
$$\epsilon_{\mathrm{F}}=\frac{3{\cdot}51}{r_s{}^2} \quad \text{and} \quad \epsilon_{\mathrm{X}}=-\frac{1{\cdot}154}{r_s} \quad \text{Rydberg units.}$$

Correlations between the positions of electrons have not been entirely neglected in the foregoing work, since the exchange energy is due to those of electrons with parallel spins. So far, however, antiparallel spin correlations have not been taken into account. No method has yet been devised for doing this in the case of real metals, but Wigner

(1934; *cf.* Seitz, 1940, p. 343) has dealt with the problem for perfectly free electrons. He uses a modified wave function of the form (9) in which the position vectors $\mathbf{s}_i$ of the negative spin electrons appear as parameters in the ψ_i of the first determinant, and the energy is then minimized. The final result is an additional energy ϵ_C per electron, called the "correlation energy," and, taking into account the increased electron density of a divalent metal, this is given by

$$\epsilon_C = -\frac{\cdot 576}{\cdot 79 r_s + 5\cdot 1} \quad \text{Rydberg units.}$$

Fig. 3.



Eigenvalue and energy curves for the metal. E_1 includes the free-electron Fermi energy. E_2 includes the Fermi energy modified according to the effective electronic mass at the zone centre. The lines *a*, *b* represent the eigenvalue and energy in a free atom as calculated by the Hartree method, and line *c* represents the total energy in a free atom, including correlation energy.

Wigner estimates that this result is accurate to within 20 per cent. Since we have already used the free-electron approximation in evaluating two of our energy terms, there is little objection to our incorporating the above expression for ϵ_C, which in any case is small, in the total energy.

Finally, then, with the approximations so far made, the total energy per electron of the crystal is

$$E_1 = \epsilon_0 + \epsilon_F + \epsilon_X + \epsilon_C$$
$$= \epsilon_0 + \frac{3\cdot 51}{r_s^2} - \frac{1\cdot 154}{r_s} - \frac{\cdot 576}{\cdot 79 r_s + 5\cdot 1} \quad \text{Rydberg units.} \quad . \quad . \quad . \quad (17)$$

E_1 is shown graphically in fig. 3, and the values of the three physical constants corresponding to this curve are given, in more conventional units, in the second line of Table III. The lattice constant, ρ, is the atomic radius at which the minimum of E_1 occurs. The compressibility, κ, referred to a single atom, is given by

$$\frac{1}{\kappa}=\frac{1}{6\pi\rho}\left(\frac{d^2E_1}{dr_s^{\,2}}\right)_{r_s=\rho}$$

The cohesive energy, S, per electron is the difference between the minimum of E_1 and the mean energy of a valence electron in the free atom (line *c* in fig. 3), and the method of calculating the latter must now be discussed. All the quantities considered here are referred to the absolute zero of temperature.

(ii) *The free atom.*

The Hartree calculation. It would be quite wrong to use the experimental value for the energy of the valence electrons in the free atom, since this would emphasize the flaws in the ion-core field which has been used in the metal. Furthermore, valence-core exchange and correlation effects have been neglected in the metal, and serious error would occur if they were included in the free atom. In fact, use of the experimental energy in the free atom would lead to a negative cohesive energy. In view of this, the wave functions and energies of the atomic valence electrons were calculated by Hartree's method, using the ion-core field derived from Yost's inner shell functions, as in the metal. The equation used was just equation (5), and the procedure exactly the same as in the calculation of ψ_0, except that r_s is now infinite. The eigenvalue was found to be $-{\cdot}456$ Rydberg units (line *a* in fig. 3), and the ϵ_0 curve must approach this value with increasing r_s. Now, in the Hartree method, the total energy of the electrons is found by subtracting their mutual potential energy from the sum of the eigenvalues. When this has been done it is found that the mean energy of a valence electron in the free atom is $-{\cdot}708$ Rydberg units (line *b* in fig. 3). In this, as in the energy calculations for the metal, the third figure is unreliable.

The correlation energy. A difficulty now arises which is not encountered in the case of monovalent metals. In the above calculation, since the two valence electrons are assumed to have opposite spins, there is no exchange energy. There will, however, be an energy term due to correlations between the electronic positions, which does not appear in the Hartree approximation. This must be taken into account, since an analogous term has been included in the energy of the metal. In this case it is not advisable to disregard both terms, since it seems likely that the correlation energy in the metal will be much greater than that in the free atom. As there is at present no method of calculating the correlation energy directly, an attempt to estimate this was made in the following manner:

The two valence electrons in magnesium spend most of their time in a region where the potential field of the ion-core is almost the same as that of the helium nucleus. It seems reasonable to assume, therefore, that the correlation energy of the valence electrons in a magnesium atom will bear approximately the same ratio to their observed energy as in the case of helium. The energy of a helium atom obtained by the Hartree method (Bethe, 1933) is $-5{\cdot}734$ Rydberg units, while the observed value is $-5{\cdot}810$. The difference of $\cdot076$ Rydberg units is due to the correlation interaction of the electrons. Thus the correlation energy in a helium atom amounts to $\cdot013$ of the observed energy. Now the observed mean energy of a valence electron in a magnesium atom is $-{\cdot}834$ Rydberg units, so that we may expect the correlation energy per electron to be approximately $-{\cdot}834\times{\cdot}013=-{\cdot}011$ Rydberg units, and this, when added to the energy obtained by the Hartree method, gives a total of $-{\cdot}719$ Rydberg units per electron (line *c* in fig. 3). It was this value which was used in computing the cohesive energy.

An alternative method which suggests itself is one similar to the preceding, but based upon the Coulomb interaction energies of the electrons rather than upon their total energies. The two methods in this case, however, lead to almost identical results. Thus, the ratio of the correlation energy to the Coulomb energy of the electrons in helium, is $-{\cdot}037$, and the Coulomb energy per valence electron in a magnesium atom is found to be $\cdot252$ Rydberg units, so that the expected correlation energy is now about $-{\cdot}252\times{\cdot}037=-{\cdot}009$ Rydberg units per electron. Although this value is practically the same as the previous one, the method of deriving it is probably better, since it depends more directly upon the electronic charge distributions, and hence might be expected to apply to the electrons in other shells, for which the ion-core field does not approximate to that of helium.

Although it seems unlikely that either of these methods of estimating the correlation energy should give results of quite the wrong size, it must be admitted that neither seems to apply to the case of beryllium. The energies of neutral Be and Be^{++} have been obtained by D. R. Hartree and W. Hartree (1935), using the methods of both Hartree and Fock, and the correlation energies obtained by comparison with experiment have been discussed by Seitz (1940, p. 248). The correlation energy of the 2*s* electrons appears to be four times as large as that estimated by the above methods. We cannot at present account for this discrepancy, but there are two reasons for suggesting that the smaller value may be nearer the truth. First, the method based on the Coulomb energies gives almost exactly the *total* correlation energy of Be, although the results for the separate electronic shells are not in accord with those of Seitz: it appears that, if the correlation energy of the 1*s* electrons were just over twice that given by Seitz there would be complete agreement. Secondly, in their work on metallic beryllium, Herring and Hill (1940) have used the experimental value of the energy of the valence electrons in the free atom, and their calculated cohesive energy is much too small. If the energy given by the Fock method had been used, together with the smaller correlation energy estimated as above, the calculated cohesive energy would have corresponded more closely with the observed.

Unless this discrepancy with regard to beryllium can be completely resolved, however, we must not ignore the possibility of a correlation energy in the free atom much larger than that used in the present work, and a calculated cohesive energy correspondingly smaller.

§4. THE EFFECTIVE MASS AT THE ZONE CENTRE.

In the previous section the Fermi energy ϵ_F has been calculated on the assumption that the valence electrons are perfectly free. We shall now attempt to correct this, at least as far as states near the centre of the Brillouin zone are concerned, using a method developed by Bardeen (1938) for the alkali metals.

According to the Bloch theory, an electronic state ψ_j is not given strictly by equation (1), but is of the form

$$\psi_j(\mathbf{r})=u_j(\mathbf{r})\exp(i\mathbf{k}_j\,.\,\mathbf{r}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (18)$$

where u_j has the periodicity of the lattice and satisfies the Schrödinger equation

$$\nabla^2u_j+2i\mathbf{k}_j\,.\,\mathbf{grad}\,u_j+(\epsilon_j-\mathbf{k}_j^2-\mathrm{V_A})u_j=0. \quad . \quad . \quad . \quad . \qquad (19)$$

To be perfectly accurate, $\mathrm{V_A}$ is no longer given by (3), but must be calculated using the wave functions (18). This refinement, however, will be neglected in the following work, since the correction arising therefrom should be insignificant, as the results of using fields A and B seem to indicate. A similar assumption was implicit in the work on sodium.

When, previously, u_j was taken to be simply ψ_0, the second term in equation (19) could be neglected to a good approximation, and this gave rise to the free-electron Fermi energy. In general, this cannot be done; but, for small $\mathbf{k}_j$, the term $2i\mathbf{k}_j\,.\,\mathbf{grad}\,u_j$ may be treated as a perturbation. In this case, following Bardeen, we shall assume that

$$u_j(r)=\psi_0(r)+i\mathbf{k}_j\,.\,\mathbf{r}\phi(r),$$

where

$$\phi(r)=\frac{\mathrm{P}(r)}{r^2}-\psi_0(r), \qquad \phi(r_s)=0,$$

and the equation for $\mathrm{P}(r)$ is

$$\frac{d^2\mathrm{P}}{dr^2}+\left(\epsilon_0-\mathrm{V_A}-\frac{2}{r^2}\right)\mathrm{P}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (20)$$

Then, to the second order in $\mathbf{k}_j$, we find

$$\epsilon_j = \epsilon_0 + \alpha \mathbf{k}_j^2, \quad \ldots\ldots\ldots \quad (21)$$

with

$$\alpha = \gamma \left[\frac{r}{P} \frac{dP}{dr} - 1 \right]_{r=r_s},$$

and

$$\gamma = \frac{4\pi}{3} r_s^3 [\psi_0(r_s)]^2.$$

The quantity α is the ratio, m/m^*, of the mass of an electron to its effective mass near the centre of the zone. If it is assumed that all the electrons have this same effective mass, the resulting Fermi energy ϵ'_F is just α times the free-electron Fermi energy ϵ_F.

TABLE II.

r_s (Bohr units)	γ	α	ϵ'_F (Rydberg units)
3·28	1·112	1·321	0·431
3·92	0·969	1·143	0·261
4·56	0·838	1·100	0·186

TABLE III.

	Lattice constant ρ (Å)	Cohesive energy S (K cal./mole)	Compressibility $10^6 \kappa$. (cm.²/kg.)
Experimental	1·77	38	3·0*
Calculated (E_1)	2·00	43	3·2
Calculated (E_2)	2·20	25	3·4

Equation (20) was solved, and α calculated, for three atomic radii covering the important part of the range; intermediate values being obtained by quadratic interpolation. The directly computed values of α and the resulting Fermi energy ϵ'_F are given in Table II. The new total energy E_2 per electron, in which all the terms except the Fermi energy are the same as in E_1, is shown in fig. 3, and the calculated physical constants are given in the last line of Table III.

The values of α corresponding to the observed lattice constant, the minimum of E_1, and the minimum of E_2, are 1·29, 1·17, and 1·11 respectively. In the case of beryllium, Herring and Hill (1940) found

---

* This is the value at room temperature: at absolute zero the compressibility will be slightly smaller than this.

α to be ·62 for the observed lattice constant. This large difference is not unexpected, however, since the corresponding values for sodium and lithium (Bardeen, 1938) are 1·07 and ·65 respectively.

§5. DISCUSSION OF RESULTS.

The calculated values of ρ, S, κ are all in fair agreement with the experimental values, but those based on the free-electron Fermi energy are uniformly better than the others. In order to assess the significance of these results correctly the effects of several approximations must be taken into account.

First, the exchange and correlation energies of the metallic electrons have been calculated using the free-electron approximation. One consequence of this is that the curves for E_1 and E_2 in fig. 3 will not, as they should, approach asymptotically the line c, representing the energy in the free atom, but the line a, representing the eigenvalue in the free atom. This is not a serious difficulty, however, since we are only concerned with a small range of values of r_s in the immediate neighbourhood of the minima of E_1 and E_2, and, in this range, the wave functions are so flat that the free-electron approximation cannot be far wrong. The possible error of 20 per cent in Wigner's formula for the correlation energy in the metal contributes an uncertainty of about ·01 Rydberg units per electron (6 K cal./mole) to the value of the cohesive energy, and it is believed that the error due to using the free-electron exchange energy will not add greatly to this. The estimate of the correlation energy in the free atom has already been discussed in §3(ii). It should be noted that an error of 1 per cent in the total energy of the metal produces an error of more than 10 per cent in the cohesive energy, and this must be taken into account when comparing the calculated with the experimental values of the latter.

Secondly, exchange and correlation interactions between the core and valence electrons have been neglected in this treatment. Gorin (1936), for potassium, and Fuchs (1935), for copper, have found such interactions to be important, but, since the ion-core of magnesium is smaller in relation to the size of the atomic sphere than for either of these metals, it is reasonable to assume that the effect will be much less in this case. The value of the cohesive energy, at least, should not be greatly affected, since the error in the energy of the metal will largely be cancelled by a similar error in the energy of the free atom.

Finally, the effect of the zone boundaries has not been taken into account. In magnesium, unlike the alkali metals, the first zone is almost full and there is a small overlap of electrons into the second zone. Within the scope of the present work we are only concerned with the effect of this upon the Fermi energy. In this respect, if Bardeen's method of calculating the effective mass at the zone centre is correct, as the results for sodium and lithium seem to indicate, it appears from Table III that the energy gaps at the zone boundaries must distort the distribution of states curve from the parabolic form in such a way that the resulting Fermi energy is not much different from that for free electrons. The calculation of the Fermi energy by Fröhlich (1936) yields a value considerably less than that for free electrons, but, owing to the approximations involved in the work, this result cannot be regarded as conclusive. Skinner (1940) finds the observed width of the soft X-ray emission band of magnesium to be 6·2 ± ·3 eV. compared with the free-electron band width of 7·3 eV., but, until the distribution of states is known more accurately, only qualitative conclusions can be drawn regarding the Fermi energy. However, in view of this evidence, and the results of the present work, it is reasonable to conclude that the Fermi energy of magnesium is not greatly in excess of that calculated on the free-electron approximation. A Fermi energy even smaller than the latter would improve the calculated value of the lattice constant, but would tend to make the value of the cohesive energy too large.

I am glad of this opportunity to express my gratitude to Professor H. Jones for many valuable discussions.

REFERENCES.

BARDEEN, J., 1938, *J. Chem. Phys.*, **6**, 367.
BETHE, H., 1933, *Handbuch der Physik*, **24/1**, 368.
FRÖHLICH, H., 1936, *Elektronentheorie der Metalle* (Berlin: Julius Springer); 1937, *Proc. Roy. Soc.* A, **158**, 97.
FUCHS, K., 1935, *Proc. Roy. Soc.* A, **151**, 585.
GORIN, E., 1936, *Physik. Z. Sowj.*, **9**, 328.
HARTREE, D. R., and HARTREE, W., 1935, *Proc. Roy. Soc.* A, **150**, 9.
HELLMANN, H., and KASSATOTSCHKIN, W., 1936, *Acta Physicochimica U.R.S.S.*, **5**, 23.
HERRING, C. C., and HILL, A. G., 1940, *Phys. Rev.*, **58**, 132.
MANNING, M. F., and KRUTTER, H. M., 1937, *Phys. Rev.*, **51**, 761.
SEITZ, F., 1940, *Modern Theory of Solids* (New York: McGraw-Hill).
SKINNER, H. W. B., 1940, *Phil. Trans. Roy. Soc. A*, **239**, 95.
WIGNER, E., 1934, *Phys. Rev.*, **46**, 1002.
WIGNER, E., and SEITZ, F., 1933, *Phys. Rev.*, **43**, 804; 1934, *Ibid.*, **46**, 509.
YOST, W. J., 1940, *Phys. Rev.*, **58**, 557.

LIV. *Nuclear Reactions Produced by Slow Negative π-Mesons.*

By M. G. K. Menon, H. Muirhead, and O. Rochat,
The H. H. Wills Physical Laboratory, University of Bristol *,

[Received April 26, 1950]

[Plates XVIII. & XIX.]

Contents

	PAGE
INTRODUCTION	584
§ 1. Statement of the problem	584
§ 2. Scope of the present experiments	585
PART I.—*General characteristics of the disintegrations and separation of the stars due to light and heavy elements*	586
§ 3. Definition of σ-stars; classification of ejected particles	586
§ 4. Identification of stars from light and heavy nuclei by the use of "sandwich" plates—Method (*a*)	588
§ 5. Separation of stars by observations on ejected α-particles and protons—Method (*b*)	590
§ 6. Separation of stars by observations on Auger electrons—Method (*c*)	592
§ 7. Unobserved disintegrations produced by π^--particles	593
§ 8. Recoil fragments	594
§ 9. The emission of energetic charged particles from σ-stars	595
PART II.—*Interaction of slow π^--mesons with heavy nuclei*	598
§ 10. General features	598
§ 11. Frequency distribution of the numbers of charged particles emitted in the disintegration of heavy nuclei	598
§ 12. The α/p ratio	601
§ 13. The energy distribution of the particles produced in the evaporation process	601
§ 14. Energy spectrum above 30 MeV	602
§ 15. Discussion	604
PART III.—*Interactions of slow π^--mesons with light nuclei*	607
§ 16. General features of the disintegrations	607
§ 17. Capture by ^{14}N	610
§ 18. Capture by ^{12}C	612
§ 19. Comparison of the disintegrations of light and heavy nuclei	616
CONCLUSIONS	616
ACKNOWLEDGMENTS	618
REFERENCES	618

Abstract.

Reactions produced by the nuclear capture of slow negative π-mesons have been studied by the photographic method. The disintegrations of the light and heavy elements of the emulsion have been distinguished by methods which are described in detail. Because of the large number of "stars" observed—more than 2500—it has been possible to make a statistical approach to the study of the disintegrations.

---

* Communicated by Professor C. F. Powell, F.R.S.

The general features of the "stars" produced by the disintegration of heavy nuclei are in satisfactory agreement with the predictions of the evaporation theory. In addition, the emission of fast, singly-charged particles has been observed, most of which must be attributed to an alternative process to that of evaporation. Preliminary results have been obtained for the frequency of emission and the energy "spectrum" of these particles. The conclusion is reached that in the initial process of capture—the so-called "primary" process—the energy corresponding to the rest-mass of the π^--meson is shared among a larger number of nucleons than has commonly been assumed hitherto.

Observations on "stars" produced by the disintegration of light nuclei indicate that there is a wide variety in the characteristics of the disintegrations which appear to be strongly dependent on the manner in which the primary distribution of energy takes place. A detailed analysis has been made of some of the disintegrations due to $^{12}_{6}C$ and $^{14}_{7}N$ nuclei, by applying the conservation laws of mass energy and momentum. In the majority of cases, the primary interaction results in the production of one or two fast neutrons and an excited residual nucleus which then disintegrates. In a few cases, however, the primary interaction involves the nucleus as a whole, resulting in its total disruption and the ejection of fast α-particles.

INTRODUCTION.

§ 1. STATEMENT OF THE PROBLEM.

AMONG the first mesons observed in photographic plates exposed to the cosmic radiation were those which, at the end of their range in the emulsion, produced nuclear disintegrations. These particles were described, phenomenologically, as σ-mesons, and subsequent work has shown that they are all, or almost all, π^--particles. A disintegration of this type is due to the capture of a meson by a nucleus; and to the distribution, among the nucleons, of the energy liberated by the disappearance of the rest-mass of the π-particle.

Whilst this view of the origin of the disintegrations produced by σ mesons appears to be beyond question, the detailed characteristics of the physical process by which the rest-energy of the meson is communicated to the nucleons have remained obscure. It appears reasonable to consider the disintegration as being produced, at least in the case of heavy nuclei, in two stages:—the "primary process", in which the π^--particle interacts with a limited number of nucleons, is followed by a general excitation of the nucleus as a result of the collisions of the fast nucleons formed in the primary process. This excitation leads to an "evaporation" of charged particles and neutrons, and the formation of the "star".

The primary interaction accompanying the nuclear capture of mesons has been studied theoretically by Bruno (1948), Marshak (1949) and other writers. The simplest example of the process would be the capture of a π^--particle by a proton, the latter being transformed into a neutron.

In order to provide for the conservation of mass-energy and momentum, such an interaction must, however, be accompanied either by the emission of a photon—the so-called radiative capture (Marshak, *loc cit.*)—or of a neutretto ; and any such neutral radiation would absorb most of the rest-energy of the meson.*

The above features of the postulated interactions of π^--particles with individual free protons, makes it necessary to assume that in the capture of mesons by heavier nuclei, more than one nucleon is involved in the primary act of absorption. For if the process of radiative capture occurred, involving a single proton, most of the rest-energy would commonly escape from the nucelus as γ-radiation or as a neutretto ; and little would be available for producing the subsequent disintegration of the nucleus. Contrary to this conclusion, it is frequently observed that the energy represented by the disintegrations produced by σ-mesons is a large fraction of the rest-energy of the meson.

These difficulties can be avoided if the primary process of capture involves two or more nucleons, for the conservation laws can then be satisfied without the assumed emission of photons or neutrettos. It has therefore been suggested (Perkins 1948, Heidmann and Leprince-Ringuet 1948) that the primary act of absorption invloves a neutron–proton pair in the nucleus ; and that, as a result, two fast neutrons of energy $\sim$70 MeV. are produced. Alternatively, it has been suggested that even more nucleons may be involved. Such a view has been adopted by Ruddlesdon and Clark (1949) in a theoretical treatment of the capture of π^--particles by helium nuclei, ${}^4_2$He.

§ 2. Scope of the Present Experiments.

2553 examples of σ-disintegrations recorded in photographic emulsions have now been examined in this laboratory, a number sufficiently great to allow a statistical analysis to be made. It has thus been possible to begin a detailed study of the nature of the processes associated with the nuclear capture of the mesons, and the subsequent nuclear explosion which commonly leads to the formation of a " star ".

One serious obstacle in making such an investigation arises from the complex nature of the emulsion, which usually makes it difficult to identify the type of nucleus involved in any particular disintegration. The first step, described in Part I. has therefore been to separate the observed stars into two groups :—those originating from the interactions of π^--mesons with the light nuclei, carbon, nitrogen and oxygen on the one hand, and those from heavy nuclei, mostly silver and bromine, on the other.

---

* *Note added in proof.*—In recent experiments at Berkeley, it has been observed that high energy γ-radiation is produced as a result of the capture of π^--mesons in liquid hydrogen. The results can be interpreted in terms of a competition between two types of transmutation corresponding to the emission of either a γ-ray of energy about 140 MeV., or the emission of a π^0-particle which subsequently decays into two γ-rays. (Panofsky 1950, Marshak and Whitman 1950.)

Heidmann and Leprince-Ringuet (1948) and Perkins (1949), have previously separated the σ-stars in a similar manner, but the statistical weight of the experimental material at their disposal was small.

The separation of the stars due to light and heavy elements has been carried out by three methods which yield results in satisfactory agreement with one another. The "prong distributions"—the relative frequencies with which the disintegrations are accompanied by the emission of different numbers of charged particles—can thus be determined separately for the two classes of stars. Examples in which the capture of the π-meson does not result in a visible "star", or in which it produces only a recoil particle of very short range, are also discussed.

In disintegrations produced by the capture of π-mesons by heavy elements, described in Part II., the "primary" process is followed by nuclear excitation which leads to the evaporation of nucleons and the formation of the star. Observations on the characteristics of these stars permit estimates to be made of the total energy involved in the evaporation, and conclusions to be drawn regarding the primary process (Perkins 1949.)

A second important source of information about the primary interaction is provided by observations on the ejection of singly-charged particles of energy greater than 30 MeV. The frequency of occurrence of these particles has been calculated by Tamor (1950), and preliminary results from experiments made with artificially produced π^--mesons have been reported by Adelman and Jones (1949), and by Cheston and Goldfarb (1950). No attempt was made by these authors, however, to consider separately the disintegrations of light and heavy nuclei.

In the case of the disintegrations of light elements, described in Part III., it is sometimes possible, as a consequence of the simpler nature of the systems involved, to make a more detailed analysis by an application of the conservation laws, and thence to draw conclusions about the nature of the primary interactions.

Part I.

General Characteristics of the Disintegrations and Separation of the Stars Due to Light and Heavy Elements.

§ 3. Definition of σ-Stars ; Classification of Ejected Particles.

In the present experiments, 1337 disintegrations due to artificially generated mesons, and 1216 due to cosmic-ray mesons have been scrutinized. The evidence is now very strong that the negative mesons produced artificially, and the σ-mesons of the cosmic radiation are of the same type. Firstly, the masses of the two types are equal within the limits of the experimental errors ; secondly, Franzinetti (1949) has shown that, in the cosmic radiation, the number of σ-mesons which are not π^--particles is very small or zero ; and thirdly, the prong distributions of the stars produced by the two types are indistinguishable. This and other evidence makes it almost certain that the naturally occurring σ-mesons are identical with the π^--particles produced artificially at Berkeley, and in what follows this will be assumed to be the case.

For the purposes of the present paper, we assume that a meson has produced a star—a nuclear disintegration— if the tracks of one or more charged particles—other than those of electrons or μ-mesons—appear to originate from the end of the trajectory of the meson in the emulsion. For each star, a preliminary classification of the ejected particles as protons, α-particles or heavy fragments has been made; and the lengths of the tracks have been measured. The identification of the particles has been based on an inspection of the grain-density and scattering characteristics of the tracks; and also, in the case of the heavy fragments due to recoil nuclei or "fission" particles, from the lengths of the tracks. With this rough method of identification most of the singly charged particles ($^1_1$H, $^2_1$H, $^3_1$H) will be classified as protons, whilst the group ascribed to α-particles will also include some heavier fragments. The term of "heavy fragment" is reserved for, (1) heavily ionizing particles producing tracks less than 5μ in length; (2) particles producing longer tracks if their ionization appears to be much greater than that produced by an α-particle; and (3) $^8_3$Li nuclei which produce the characteristic "hammer" tracks.

TABLE I.

Element	Composition of normal emulsion		Composition of gelatin	
	In gms. per cc.	Atomic composition	Percentage composition	Atomic composition
Silver	1·85	30·9	—	—
Bromine	1·34	30·3	—	—
Iodine	0·052	0·7	—	—
Sulphur	0·010	0·8	—	—
Carbon	0·27	49·0	43·8	49·7
Hydrogen	0·056	100·0	7·4	100·0
Oxygen	0·27	26·7	33·7	28·7
Nitrogen	0·067	10·5	15·2	14·7

§ 4. THE IDENTIFICATION OF STARS FROM LIGHT AND HEAVY NUCLEI BY THE USE OF "SANDWICH" PLATES—METHOD (*a*).

In these experiments, we have used gelatine "sandwich" emulsions exposed to π^--mesons produced in the Berkeley cyclotron. The emulsions were composed of layers of pure gelatine, 5μ thick, sandwiched between layers, 30μ thick, of Ilford C2 emulsion loaded with boron; the composition of the gelatine employed in given in Table I. (col. 4). It can be seen that the atomic composition of the gelatine in the sandwich is similar to that in the normal nuclear emulsions (cols. 3 and 5, Table I.).

TABLE II.

Number of charged particles ejected per star	Phenomenological classification of the star	Number of stars (observed)	Number of stars (after correction for loss)
1	1p	5	4
	1α	5	5
	1f	3	6
2	2p	1	1
	2α	3	3
	1α, 1p	5	4
	1p, 1f	1	2
	2f	1	1
3	3α	1	1
	2α, 1p	10	10
	1α, 2p	4	4
	1α, 1p, 1f	1	2
4	2α, 2p	1	1

Of the 600 stars observed in these plates, 41 were produced in the gelatine, and therefore must certainly be attributed to the disintegration of light elements. Photo-micrographs of three typical examples of these events are reproduced in Pl. XVIII. (*a*), (*b*), (*c*). Table II. gives details of the 41 stars originating in the gelatine; the letters "p", "α", "f", refer to protons, α-particles and heavy fragments respectivley.

We have also observed 38 mesons which appeared to stop in the gelatine and did not give rise to any visible disintegration.* Some of the particles emitted from the disintegrations will not be observed either through failing to emerge from the gelatine, or as a result of entering the emulsion at a considerable distance from the parent star, so that the association of the track with the meson is uncertain. Taking account of these effects and of the smaller stopping power of the gelatine, simple geometrical considerations suggest that the probability of "loss" is less than 0·1, for α-particles and protons of energy greater than 3 MeV. and 1 MeV. respectively. For the shorter tracks due to heavy fragments, the probability of loss is greater and approximate calculations give a value of 0·5. The appropriate corrections are included in the fourth column of Table II.

---

* We cannot exclude the possibility that some of these mesons may have stopped in the emulsion, just before entering, or immediately after leaving the gelatine layer.

The experimental distribution in prong number for σ-stars, in Ilford C2 emulsion and gelatine respectively, is given in Table III. Using this experimental material, the separation of the "stars" produced in light and heavy elements of the normal emulsion has been carried out in the following manner :

TABLE III.

No. of prongs	1137 stars by π^--mesons from cyclotron (percentage)	665 stars by π^--mesons from cosmic radiation (percentage)	Total no. of stars in the emulsion (percentage)	44 stars produced in gelatine (percentage)
1	$33{\cdot}7\pm1{\cdot}7$	$34{\cdot}7\pm2{\cdot}3$	$33{\cdot}9\pm1{\cdot}4$	34 ± 9
2	$33{\cdot}1\pm1{\cdot}7$	$28{\cdot}1\pm2{\cdot}0$	$31{\cdot}3\pm1{\cdot}3$	25 ± 8
3	$22{\cdot}1\pm1{\cdot}4$	$21{\cdot}4\pm1{\cdot}8$	$21{\cdot}8\pm1{\cdot}1$	38 ± 9
4	$9{\cdot}3\pm1{\cdot}0$	$13{\cdot}5\pm1{\cdot}4$	$10{\cdot}9\pm0{\cdot}8$	2 ± 2
5	$1{\cdot}7\pm0{\cdot}4$	$2{\cdot}5\pm0{\cdot}6$	$1{\cdot}9\pm0{\cdot}3$	—
6	—	$0{\cdot}3\pm0{\cdot}2$	$0{\cdot}1\pm0{\cdot}1$	—

It has been assumed that the "three-prong" stars which appear in the normal emulsion, and which are of the same type as those observed in the gelatine (Table II.), are the products of the interaction of the π^--mesons with light nuclei. In the gelatine, the proportion of such stars with three prongs is $17/44=0{\cdot}38$. On the other hand, in the emulsion, 71 three-prong stars of the same type were found in a sample of 460 stars. The proportion is therefore $71/460x$, where x represents the fraction of the 460 stars which must be attributed to light nuclei. Assuming that the two frequencies are equal, it is found that $x=0{\cdot}40\pm0{\cdot}12$.

Using this value, one can deduce, from the observed prong distributions in normal emulsion and gelatine, the corresponding distribution for the stars produced in heavy nuclei only. The results, normalized to 100 σ-stars in the emulsion, are given in Table IV.

§ 5. SEPARATION OF STARS BY OBSERVATIONS ON EJECTED α-PARTICLES AND PROTONS—METHOD (*b*).

An alternative method of estimating whether a star is due to a light or heavy element, employed by Heidmann and Leprince-Ringuet (1948), is based on the assumption that an α-particle with an energy less than that corresponding to the potential barrier of bromine can only have emerged from a light nucleus. At first sight, it might appear that some caution should be observed in applying the method, for the production of short-lived radioactive heavy nuclei might permit the ejection of α-particles of lower energy. Further, some of the tracks, assumed to be produced by α-particles, may actually represent nuclear fragments of greater charge and mass.

TABLE IV.

No. of prongs	Method (*a*)		Method (*b*)		Method (*c*)		Mean values		
	L	H	L	H		H	L	H	L+H
1	13±4	21	10±1·5	24±4		35±5	11	27	37
2	10±3	20	15±2	18±3		18±3	13	18	31
3	15±4	7	15±2	5·5±1·5		1±1	15	4	20
4	2±1	10	8±1·5	3·5±1·0		—	6	4	10
5	—	2	1·5±0·8	0		—	1	1	2
6	—	—	—	0·5±0·5		—	—	—	—
Totals	40	60	49	51		(54)	46	54	100

Percentages of σ-stars of different types produced in light elements (L) and heavy elements (H) as determined by different methods. The absolute values for method (*c*) have been deduced from the observations, assuming that 54 per cent of the σ-stars originate in heavy nuclei. 21 per cent of the σ-stars have associated electrons.

Strong support for the "potential-barrier" method is given, however, by the observed distribution in energy of the α-particles emerging from those σ-stars in which a single α-particle track is the only visible product of the disintegration. Such an energy distribution, deduced from observations on 146 tracks, is shown in fig. 1, suitable geometrical corrections having been made to allow for the "loss" of tracks of great range.

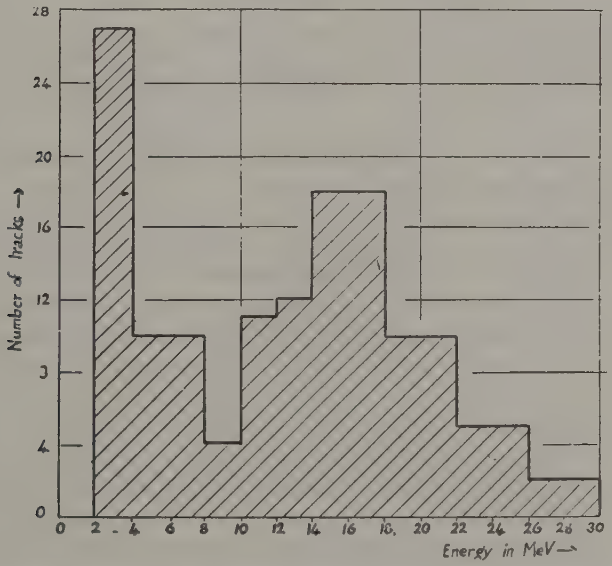
The results shown in fig. 1 suggest that the single tracks identified as due to α-particles, fall into two groups. We shall show in a later section that many of the tracks which have been attributed to low energy α-particles, must in fact be due to the recoil of a light nucleus. The α-particles in the high energy group, which can be identified with much more confidence, must be attributed to the disintegration of heavy nuclei; for the emergence of a high energy α-particle from a light element would normally be accompanied by a detectable recoil of the residual nucleus.

It can be seen from fig. 1 that the point of separation of the two groups of α-particles is situated at ∼9 MeV. Perkins (1949) by a similar method obtained a value of ∼10 MeV.

On the basis of the above result, it has been assumed that a partial separation of the "stars" can be made by considering the range of the α-particle tracks, irrespective of the "prong-number" of the star. The division has been made in the following manner :—

If one or more α-tracks in a star have ranges corresponding to a kinetic energy of less than 9 MeV., the star has been attributed to a light nucleus. Further, in the cases where no α-particles are emitted, the occurrence of a proton of energy less than 4 MeV. has been taken to indicate disintegration of a light nucleus ; the possibility of proton decay of heavy nuclei can hardly lead to serious errors. Among 500 stars which were examined, it was found that 214 could be attributed to the disintegration of light nuclei, and 162 to heavy nuclei. Among the others, 32 were composed of short range fragments only. In 92 cases all the particles left the emulsion after traversing a distance smaller than 50μ (for an α-particle) or 120μ (for protons). These distances correspond to the limits of energy considered in our above criterion.

Fig. 1.



Histogram showing the distribution in energy of α-particles from 147 single-prong stars.

It is reasonable to suppose that most of the 32 disintegrations in which heavy fragments only were emitted, are due to the break up of light nuclei. Thus it will be shown in a later section that the maximum energy that can be imparted to a heavy recoil fragment coming from a disintegrating silver or bromine nucleus is 6 MeV. The average energy of such a fragment will be much smaller—only ∼0·7 MeV. Such fragments will, in the majority of cases, be unable to produce an observable track. The 92 stars in which the lengths of the tracks in the emulsion were too short for the above criterion to be applied have all been attributed to heavy nuclei, since the prong distribution for these 92 cases was very similar to that for the 162 cases which could be definitely classified.

The fraction of σ-stars attributed to light nuclei by this method is 0·49, a result in satisfactory agreement with the data based on the stars in gelatine. The corresponding prong distribution for light and heavy elements normalized to 100 stars in the emulsion is included in Table IV.

§ 6. SEPARATION OF STARS BY OBSERVATIONS ON AUGER ELECTRONS—METHOD (*c*).

A third method for distinguishing the stars produced in light and heavy elements is based on the observation of tracks of slow electrons emerging from some of the σ-stars. A typical example is shown in Pl. XIX. (*a*). Although all these electrons are not necessarily produced in the same type of process, the evidence is very strong that they are all associated with the capture of π^--particles by heavy nuclei.

Most of the electrons probably arise in the process of atomic capture of the negative mesons. The energy released in the transition of the meson from one level to another can be transferred to an atomic electron—a form of Auger effect. The recent work of the Brussels group (Cosyns *et al*. 1949) has shown that this process occurs in the capture of μ^--mesons, and that slow electrons are rarely associated with the mesons which suffer direct β-decay. The last point strongly suggests that, in nuclear emulsions, the Auger electrons only appear during the capture of μ^--mesons by silver and bromine nuclei, for it is well known that with such heavy elements, the meson commonly interacts with the nucleus before suffering β-decay. The observations of the Brussels group on the energy spectrum of the slow electrons provides further support for this conclusion; and it is also confirmed by the work of Franzinetti (1950), who finds that no slow electrons accompany the decay of μ^--particles.

The slow electrons under consideration might also be attributed to the internal conversion of γ-rays, emitted after the capture of negative mesons by nuclei. In this case, however, the emission of a slow electron from the point where the nuclear disintegration occurred would only take place if a residual radiaoctive nucleus remained after the disintegration. There is good evidence, however, for the view that, following the interaction of a π^--meson with a light nucleus, the whole nucleus commonly disintegrates, and that no residual nucleus remains. A third possible origin of the slow electrons could be the emission of a β-decay particle as the final stage of the "cooling" of a nucleus excited by the capture of a π^--meson. Here again, however, the process is restricted to heavy nuclei.

It is reasonable to suppose that the process of atomic capture of π^- and μ^--mesons will differ only as a result of the difference in mass of the two particles, which will be manifest in a slightly greater energy of any emitted slow electrons in the case of the heavier π^--particles. 400 σ-stars, formed by π^--mesons of the cosmic radiation in Ilford G5 and Kodak NT4 emulsions were therefore examined for the presence of slow electrons. The slow-electron back-ground in an "electron-sensitive" emulsion is commonly high. Through a random coincidence, the tracks of unassociated slow electrons will therefore sometimes appear to diverge from the point where a meson stops in the emulsion. A correction for this effect was made by examining the point of the decay of 133 π-mesons for the presence of similar unrelated electrons. Three apparently associated electron tracks were found. The plates used were the same as those employed for the examination of σ-stars, and the slow electrons may have been due either to random coincidences, or to δ-rays produced by μ-mesons near their point of origin. The last feature would correspond very roughly to the chance production of a δ-ray by one of the products of disintegration of a σ-star.

The relative frequencies with which stars due to heavy nuclei, as identified by the above method, have different numbers of prongs, are shown in Table IV. The method does not give absolute values because the probability of the Auger transitions is not known, but the numbers given in Table IV. have been calculated on the assumption that 54 per cent of the σ-stars are due to heavy nuclei.

The final conclusions regarding the prong distributions from light and heavy elements are included in Table IV. They were obtained by a combination of the values given by the three methods (*a*), (*b*) and (*c*). The errors indicated throughout are the standard deviations. On the basis of the results obtained with methods (*a*) and (*b*), 54 ± 5 per cent of the stars observed have been attributed to heavy nuclei, and 46 ± 5 per cent to light nuclei.

Of the 57 single-prong stars accompanied by Auger electrons, 11 were identified as 1α, and 32 as 1p stars. In the remaining 14, the length of path in the emulsion of the secondary particle was not sufficiently great for an identification to be made. The value of the α/p ratio for heavy nuclei corresponding to these figures is 0·3, a result which may be attributed to the different heights of the potential barriers for protons and α-particles.

It is to be expected that the α/p ratio for light nuclei should be greater than that for heavy nuclei. In accordance with this expectation, the ratio of the number of α-particles and heavy fragments to that of protons was found to be $1{\cdot}55 \pm 0{\cdot}2$ for the stars in gelatine; and for the 246 stars identified by the potential barrier method as due to the disintegration of light nuclei, the α/p ratio was found to be $273/271 = 1$.

§ 7. UNOBSERVED DISINTEGRATIONS PRODUCED BY π^--PARTICLES.

Many of the π^--mesons, on being brought to rest in the emulsion, do not produce observable disintegrations (Occhialini and Powell 1948).

Thus Adelman and Jones (1949) have observed that 27±2 per cent of the π^--mesons do not produce a visible star. Our results from similar data give a value of 28±2 per cent.

We have observed that many of these "star-less" π^--mesons, at the end of their range, produce a "blob"—a rudimentary track of very restricted range, apparently due to a heavily ionizing particle. Adelman has informed us in a private communication that he has made similar observations. A typical example is shown in Pl. XVIII. (*d*). Whilst it is premature to attempt a final explanation of these events, the observations of Franzinetti (1950) strongly suggest that they are associated with the capture of negative mesons by silver and bromine nuclei. In an examination of 100 examples of such starless mesons, in plates exposed in the Berkeley cyclotron, 40 have been found to have blobs associated with them. We regard this figure as representing a lower limit to the proportion of "starless" mesons captured in heavy nuclei.

It is possible to make an estimate of that fraction of the starless mesons which corresponds to cases in which the π^--particles have been captured by hydrogen :—Our results show that out of 100 σ-stars observed in the emulsion, 54±5 occur in silver and bromine, and 46±5 in light nuclei. The number of starless mesons corresponding to 100 mesons producing stars is 39. If we assume, (1), that the capture of a meson by the nuclei of C, N, O always results in a visible "star" (see later section regarding cases in which only a recoil is produced) ; and, (2), that the probability of capture by different kinds of nuclei is proportional to the atomic charge (Fermi and Teller, 1947), then the 45 stars produced in C, N, O nuclei will be accompanied by 8 cases in which the meson is captured by a hydrogen atom. This result suggests that 80 per cent of the π^--mesons which stop in the emulsion without producing an observable disintegration correspond to cases of capture by heavy nuclei in which neutrons alone are "evaporated".

§ (8). Recoil Fragments.

The tracks of heavily ionizing particles of short range have been observed in 27±2 per cent of the disintegrations with which one or more prongs are associated. It is important to decide whether these are due to heavy nuclei, recoiling in order to provide for the conservation of momentum in the disintegrations.

The kinetic energy of such a heavy fragment will result from a process in which energy of such a heavy fragment will result from a process in which $n+1$ particles (including the recoil fragment itself) share a total energy E in such a way that the total momentum is zero. If A is the total mass of the nucleus, M the mass of the recoil fragment, and E the total available energy, then the maximum energy of the recoil nucleus, $E_r(\max)$, is $E(A-M)/A$. E is equal to $\mu c^2 - \sum_{i=1}^{i=n} B_i$ where μc^2 represents the energy equivalent to the rest-mass of the meson, and B_i the binding energy of the emitted particles. If it is assumed, for simplicity, that only individual nucleons are emitted, then we may write $\sum_{i=1}^{i=n} B_i = 8n$ MeV. for nuclei of intermediate mass (n being the number of nucleons emitted). We therefore obtain $E_r(\text{max}) = (140-8n)n/A$.

The maximum value of the above expression for E_r(max) occurs for $n=9$ and gives a value of 6 MeV. for $A=100$. This corresponds to a very improbable case in which the nine evaporated particles are emitted in the same direction. The average energy of a recoil fragment from a heavy nucleus will be one-ninth of the maximum. These considerations make it unlikely that any track longer than 3μ is due to the recoil of a heavy nucleus, and justify the previous assumption that the stars containing only heavy fragments are due to the disintegration of light nuclei.

Harding (1949) has shown that among the stars produced in plates exposed to the cosmic radiation by fast neutrons, it is possible to identify short thick tracks due to the recoil of heavy nuclei. In such cases, however, an important part of the energy of the recoil fragment is due to the motion of the centre of mass (c.m) of the interacting particles, whereas in the nuclear capture of π^--mesons the centre of mass is at rest.

§ 9. THE EMISSION OF ENERGETIC CHARGED PARTICLES FROM σ-STARS.

The charged particles emitted as a result of the nuclear capture of π^--particles may be divided into two categories : (*a*), low energy nucleons or fragments which result from the evaporation processes in heavy nuclei or from the disintegration of light nuclei ; and ,(*b*), fast particles arising directly from the primary process (Tamor 1950). The production of thick " electron-sensitive " plates has made it possible to study these fast particles. We have examined a total of 361 σ-stars recorded in Ilford G5 emulsion 400μ thick, exposed to the cosmic radiation at an altitude of 70,000 feet, and determined the number of tracks with a grain density less than 150 grains/100μ.

In the particular conditions of development of the emulsions, these tracks are due to particles with the following characteristics :—protons with energy greater than 30 MeV.; deuterons with energy greater than 60 MeV. ; and tritons with energy greater than 90 MeV. Only tracks of which the horizontal projection was greater than 200μ were accepted for measurement, in order, (1), to maintain statistical accuracy in the determination of the grain density ; and, (2), to distinguish the fast particles giving rise to the single prong of some σ-stars from examples of $\pi \rightarrow \mu$ decay. A simple geometrical analysis allows the total number of particles in the stated energy interval, associated with the stars examined, to be determined from the observed numbers of tracks of range greater than 200μ, and the results are shown in Table V. The correctness of this procedure has been checked by determining approximate values of the grain-density in all the tracks, irrespective of their length, and the results thus obtained are included in the Table V. It will be seen that there is satisfactory agreement between the results obtained by the two methods.

The separation of the results given in Table V. into those corresponding to light and heavy nuclei, respectively, has been carried out by the potential barrier method described previously, and from evidence based on the observation of slow electrons associated with the star. The results indicate that $4{\cdot}9 \pm 1{\cdot}5$ per cent of the stars observed in light nuclei and $21{\cdot}4 \pm 2{\cdot}5$ per cent of the stars observed in heavy nuclei have associated fast particles of charge $|e|$; and that $11{\cdot}9 \pm 1{\cdot}4$ per cent of all the disintegrations are accompanied by the emission of fast protons or heavier hydrogen nuclei with energy greater than 30 MeV. This figure does not take into account the "starless" mesons which form 28 ± 2 per cent of the

TABLE V.

Stars examined		Observed no. of particles (G.D.$<$150 gr./100 μ ; range in emulsion $>$200 μ)	Calculated no. of particles after geometrical correction	Observed no. of fast particles, including those with range in the emulsion $<$200 μ
In heavy nuclei :	159	21	34	33
In light nuclei :	202	6	10	10
Total :	361	27	44	43

TABLE VI.

Range in μ	Particle	Energy of particle in MeV.*	No. of prongs	Probable nucleus
$>$1150	p	38 ± 5	1	Ag or Br
4070	p	33 ± 1	1	Ag or Br
$>$1500	p	37 ± 4	1	Ag or Br
$>$1200	p	60 ± 8	1	Ag or Br
$>$2550	p	56 ± 4	4	Oxygen
$>$1000	p	32 ± 5	2	Ag or Br
$>$3760	p	53 ± 4	4	Oxygen
$>$1700	p	51 ± 5	1	Ag or Br
$>$3500	p	82 ± 6	1	Ag or Br
$>$1200	p	31 ± 5	1	Ag or Br
$>$5900	p	62 ± 4	1	Ag or Br
5250	d	50 ± 2	1	Ag or Br

* In this table the errors indicated are "probable errors"

total number of mesons captured in the emulsion. If these are included, the results therefore indicate that $9{\cdot}5 \pm 1{\cdot}4$ per cent of all the mesons captured by atoms result in the emission of fast singly-charged particles.

Cheston and Goldfarb (1950) have reported that 8 ± 2 per cent of the particles emitted from σ-disintegrations are protons with energies between 30 and 70 MeV. The present observations, in which all the particles arising fron 361 stars have been scrutinized, lead to the conclusion that $5{\cdot}7 \pm 1{\cdot}4$ per cent of the ejected particles are protons with energy greater than 30 MeV.

In some cases, the length of the trajectory of a fast particle in the emulsion is greater than 1000μ. It is then possible to determine both the grain-density and Coulomb scattering of the track, and thus to measure the mass and energy of the corresponding particle. The results are given in Table VI.

Tamor (1950) has calculated the probability of the emission of these fast particles, and his theoretical results are shown in Table VII. In making the calculations, he considers two alternative mechanisms for the capture process. In the first, it is assumed that the primary interaction of the π^--meson takes place with a neutron–proton pair. Two fast neutrons are thus produced which travel in opposite directions, each with an energy of 70 MeV. The second model assumes that the primary interaction is with an α-particle group, which disintegrates with the emission of a 90 MeV. neutron and a 30 MeV. triton. These " primary " neutrons are assumed to create fast charged " secondary " particles in collisions with other nucleons of the nucleus. In all these calculations, charge exchange is assumed to occur in half the neutron–proton collisions.

Table VII.

	Model	Nitrogen	Silver
No. of fast protons (E>30 MeV.) per 100 mesons absorbed	I—neutron-neutron	48	24
	II—neutron-triton	12	13

Whilst the present experimental data is not of sufficient statistical weight to allow any final conclusions to be drawn, it is possible to indicate general tendencies, and to make a comparison with Tamor's theoretical predictions.

The observed frequency of occurrence of fast protons associated with σ-stars ($9{\cdot}5$ per cent) agrees very well with the results based on Tamor's model (II.). This may be a coincidence, however, for there is a large disparity between the observed and calculated ratios of the frequency of occurrence of fast particles emitted in the disintegration of light and heavy nuclei, respectively. On the other hand, the neutron–neutron

Model (I.) suggests a much greater proportion of fast protons than is observed ; in particular, the proportion of 48 per cent obtained from this model, in the case of light nuclei, appears to be much too great. Although the statistical weight of the present observations is small, they strongly suggest that the ratio of the number of fast particles, arising from the disintegration of light nuclei, to the corresponding figure from heavy nuclei, is considerably less than 1, whereas Tamor's calculations, based on Models I. and II. give values of 2 and 1, respectively. Further comments on these questions are made in a later section.

PART II.

Interaction of Slow π^--Mesons with Heavy Nuclei.

§ 10. GENERAL FEATURES.

The problem of the interaction of slow π^--mesons with heavy nuclei (Ag and Br) has received more extensive theoretical treatment than the corresponding problem with light nuclei (Perkins 1948, Fujimoto and colleagues 1949). In these theoretical calculatons, it has generally been assumed that the primary process leads to the excitation of the nucleus, which subsequently " evaporates ". By assuming particular mechanisms for this process, it is possible to calculate the initial excitation energy, a quantity upon which the following features of the subsequent nuclear disintegration depend :—

(*a*) the prong distribution for the stars and the average number of ejected charged particles per star ;

(*b*) the α/p ratio ; and

(*c*) the distribution in energy of the emitted particles of charge $|e|$. Fujimoto and Yamaguchi (1949), and Le Couteur (1950), have treated these problems in considerable detail. Further, investigations on :—

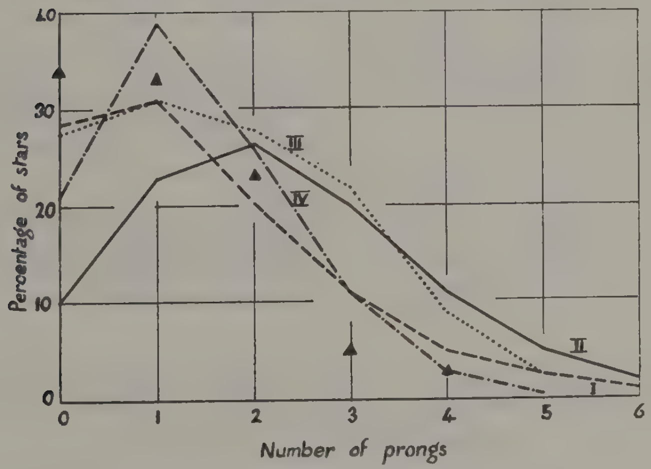
(*d*) the distribution in energy of the fastest particles of charge $|e|$— already briefly discussed in the preceding section, enable tentative conclusions to be drawn regarding the mechanism of the " primary " excitation.

§ 11. FREQUENCY DISTRIBUTION OF THE NUMBERS OF CHARGED PARTICLES EMITTED IN THE DISINTEGRATION OF HEAVY NUCLEI.

In determining the " prong " distribution for the disintegrations of heavy nuclei, the relative frequencies with which the disintegrations result in the ejection of different numbers of charged particles, account must be taken of that fraction of the " starless " mesons which correspond to capture by Ag and Br. Two methods of estimating this quantity have been described in a previous section. The results obtained show that 54 ± 5 per cent of the visible disintegrations in the emulsion occur in heavy nuclei ; and that every 54 of these events are accompanied by 31 cases* in which the meson is captured by a Ag or Br nucleus without producing a star. This estimate has been employed in determining the distribution shown below, which corresponds to 100 mesons captured by Ag or Br nuclei :—

0–prong	36	(37)
1–prong	33±2	(54)
2–prong	22±1	(6)
3–prong	5±1	(2)
4–prong	5±2	(1)

Fig. 2.



Theoretical prong distributions for heavy nuclei.

Curve I. — — — — Fujimoto and colleagues.
Curve II. ———— Fujimoto and colleagues.
Curve III. .... Clementel and Puppi.
Curve IV. –·–·–·–·– Authors.

The experimental values are shown thus : ▲

The bracketed values are percentages deduced by Perkins from an analysis of 128 stars. The average number of prongs per disintegration is 1·1±·15 (0·78). It may be pointed out that the value for the frequency of occurrence of starless mesons, captured by heavy nuclei, is less accurate than the relative proportions of stars with 1, 2 and 3 prongs.

---

* Thus 28 per cent of the π^--mesons fail to produce a recognizable disintegration ; *i. e.* there are 39 starless mesons for every 100 σ-stars of which 54 are due to heavy nuclei. But of the 39 starless mesons, 8 are estimated to be due to interactions with hydrogen nuclei.

The present results appear to indicate a higher value of the excitation energy than that estimated by Perkins.

Fig. 2 shows the prong distributions predicted from theoretical considerations, the methods employed being indicated below.

Curve I.	Fujimoto and colleagues (1949).	Average initial excitation : 91 MeV. Average number of prongs : 1·4. Poisson's fluctuation law. Fermi-gas model.
Curve II.	Fujimoto and colleagues (1949).	Excitation energy : 140 MeV. Average number of prongs : 2·2. Poisson's fluctuation law. Fermi-gas model.
Curve III.	Clementel and Puppi (1949).	Excitation energy : 140 MeV. Average number of prongs : 1·2. Poisson's fluctuation law. Liquid-drop model (restricted to volume waves).
Curve IV.	Authors.	Excitation energy : 100 MeV. Average number of prongs : 1·2. Binomial distribution law. Fermi-gas model.

We have calculated curve IV. using a slight modification of Fujimoto's curve I., and a binomial distribution law. The average number of nucleons ejected being 7, the probability of observing ν protons is

$$\mathrm{P}(\nu)=\binom{7}{\nu}\cdot(a)^{\nu}\cdot(1-a)^{7-\nu},$$

where "a" is the probability for an ejected nucleon to be a proton. We have considered the fact that we are dealing here with a Poisson rather than a Bernouillian series, since the probability of emitting a proton changes during the evaporation. The probabilities have been determined using seven successively smaller values of the excitation energy, but this refinement has little effect on the calculated distribution.

Curve III., based on the liquid drop model, gives good agreement between the observed and calculated values of the average number of prongs, but it predicts a distribution in the energy of the emitted particles which is widely different from that observed. We therefore regard the model as unsatisfactory for the description of these processes.

In all the calculations referred to above, one important factor has been neglected :—viz., the reduction of the charge of the nucleus by one unit as a result of the absorption of the meson. This produces both a change in the binding energy and an excess of neutrons in the excited nucleus. It should therefore have the effect of increasing the relative yield of neutrons at the beginning of the evaporation. For this reason, all the theoretical calculations probably tend to underestimate the proportion of "starless" mesons.*

§ 12. The α/p Ratio.

The ratio of the number of particles with charge $|2e|$, emitted from heavy nuclei, to that with charge $|e|$, has been determined in the following manner:—

The total number of "protons" and "α-particles" has been deduced from observations on the 1-prong stars among a group of 780 recorded in C2 emulsion, making suitable geometrical corrections for "loss". The histogram showing the energy distribution of single α's (fig. 1) has been "cut off" below 9 MeV., since the particles of lower energy are believed to arise from light nuclei. The penetrability of the potential barrier has little effect on the results, for excitation energies in the range under consideration, and few of the short range α-particles can have been due to heavy nuclei.

Using this method, 58 single α-particles and 193 protons were found among the 1-prong stars. The corresponding value of the α/p ratio is $0{\cdot}3 \pm 0{\cdot}04$, identical with Perkins' estimate of 0·3. These values also agree well with Le Couteur's theory for an excitation energy of 100 MeV. A previous estimate, in which the stars from heavy nuclei were identified by the observation of slow electrons, as described in an earlier section, (§ 6), also gave a value of 0·3.

§ 13. The Energy Distribution of the Particles produced in the Evaporation Process.

The distribution in energy of particles of charge $|e|$, emitted from the 1- and 2-prong stars attributed to heavy nuclei, has been determined assuming that all the particles are protons. For this purpose, the range was measured of all of these particles which were brought to rest in the emulsion. The resulting distribution in range was then corrected for "loss", and transformed to a distribution in energy. The result thus obtained, based on measurements on 309 tracks found in a group of 1430 stars recorded in Ilford C2 and G5 emulsions, is shown in fig. 3.

---

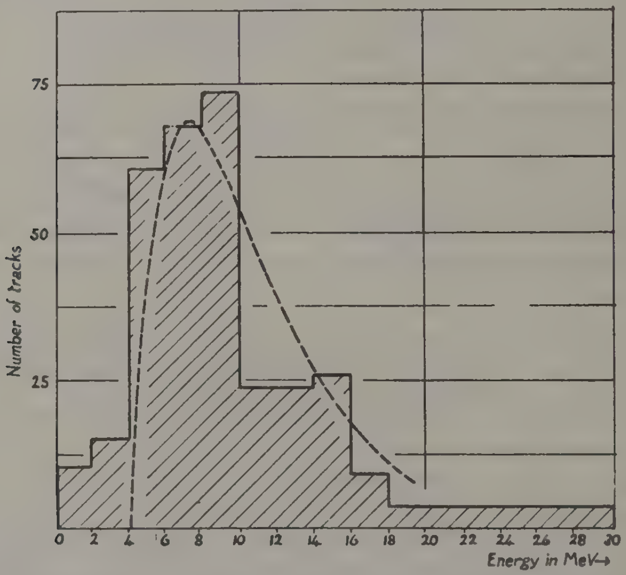
* The "heavy" nuclei considered here are those of silver and bromine. For the nuclei of lead, the average number of prongs resulting from the capture of a π-meson should be even smaller than that for silver, since the potential barrier for lead is about 10 MeV. A similar conclusion applies to the disintegrations produced by fast mesons; namely, that the average number of the secondary particles produced in lead will be smaller than that observed in the emulsion. These considerations may have a bearing on the measurements of the mean-free-path between collisions of the particles produced in penetrating showers and observed in cloud chambers crossed by lead plates. A large fraction of the disintegrations, produced in the lead plates of the thickness commonly employed, will be undetected. This will tend to lead to an overestimate of the value of the mean-free-path.

The average energy per particle—of all the particles of energy less than 30 MeV.—is 9·6 MeV. Perkins had previously obtained a value of 7·6 MeV. Using the Weisskopf evaporation theory, and the relation between excitation energy and density of states given by the Fermi gas model, the calculated distribution in energy is obtained as a relation of the form

$$P(E)\,dE=(E-4)\exp\left[-{\cdot}316(E+8)/\sqrt{(U)}\right], \quad . \; . \; . \; . \; . \quad (1)$$

where U is the excitation energy of the nucleus. With a suitable value of the parameter, U, an expression of this form is found to give a satisfactory fit with the experimental results. The value of U, thus obtained, is 109±20 MeV., the errors being those associated with statistical fluctuations only. The above treatment is very approximate, and lack of precise knowledge of the magnitude of the potential barrier may introduce systematic errors up to 30 per cent ; but it gives an estimate of the order of magnitude of the excitation energy.

Fig. 3.



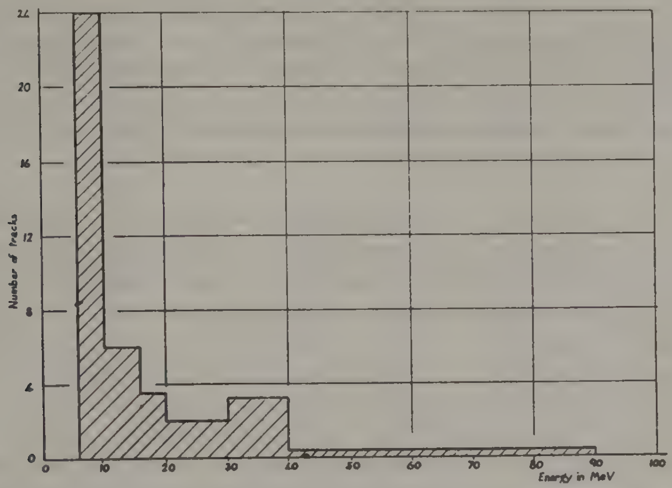
Histogram showing energy distribution of protons (E< 30 MeV.) in 1- and 2-prong σ-stars. Dotted curve is the theoretical distribution.

§ 14. Energy Spectrum above 30 MeV.

Measurements on "fast protons" from σ-stars have been described in an earlier section—§ 9. Here we shall consider the energy spectrum of those emitted from heavy nuclei.

Fig. 4 shows the distribution in energy of all the particles of charge $|e|$ observed in an examination of 146 one-prong stars originating from heavy nuclei. The presence of a considerable proportion of protons of energy greater than 30 MeV. cannot be accounted for as due to a "tail" in a Maxwellian distribution, and they must be attributed to an alternative process other than that of evaporation. The distribution in the values of the energy of the fast protons ejected from heavy nuclei is shown in Table VIII. A possible explanation of the production of these energetic protons is that they are projected from nuclei in collisions with neutrons of energy $\sim$70–90 MeV., the latter arising in the primary process.

Fig. 4.



Energy distribution of all protons ($0 < E < 90$ MeV.) from 146 single-prong σ-stars in Ilford G5 emulsion. Allowance has been made for α particles of range $> 230\,\mu$ by considering the scattering characteristics along the tracks. It can be seen from fig. 1 that in any case this proportion is negligible.

TABLE VIII.

Energy range in MeV.	30–40	40–50	50–60	60–70	70–80	80–90
Number of fast protons	14	—	3	2	—	2

Approximate estimates have been made of the ratio of the number of fast protons with energies in the interval $30 < E < 50$, to the number with energy greater than 50 MeV. that would be expected to occur if the protons are indeed due to the impact of fast neutrons, with energy $\sim$70 MeV., created in the primary process. The estimates have been based on calculations of the distribution in energy of secondary nucleons produced by the passage through nuclear matter of neutrons of about this energy, made by Goldberger (1948) and Heisenberg (1937). Both these calculations, and the experimental results of Tracy and Powell (1949) on the bombardment of carbon by 90 MeV. neutrons, give values significantly smaller than 1 for the above ratio, whereas the observed ratio is about 2. It appears therefore that most of the fast protons cannot be regarded as "knock-on" particles, produced by collision with such energetic "primary" neutrons.

§15. Discussion.

It will be convenient to summarize here the evidence obtained from the present study of the disintegrations of heavy nuclei and its bearing on the nature of the "primary" interaction responsible for the absorption of the meson.

The energy spectrum, prong distribution and α/p ratio of the stars produced in heavy nuclei, all support the view that the average excitation energy is of the order of 100 MeV. That it is, in general, less than the rest-energy of the π-meson, is clearly shown by the disagreement between the experimental results and the theoretical predictions—both for the prong distribution and the average number of prongs—based upon a value of 140 MeV. for the excitation energy. (Curves II. and III., fig. 2.). The present observations, therefore, give support to an earlier conclusion reached by Perkins and several other workers. It follows that, on the average, about a quarter of the rest-energy of the meson is lost to the nucleus by the emission of fast particles before the general excitation, and the associated process of evaporation has been established.

The observed value of the excitation energy, $\sim$100 MeV., is not widely different from that calculated on the assumption that the primary interaction takes place with a neutron–proton pair. Thus, assuming that the mean free path for collision of nucleons in nuclear matter is 4×10^{-15} . E cm., where E is the energy of the particle in MeV., Perkins (1949) has obtained a value of 80 MeV. for the average value of the excitation energy. It has been shown by Tamor (1950), however, that this type of interaction should result in a much greater proportion of fast protons, of energy greater than 30 MeV., than is actually observed :—24 per cent from heavy nuclei and 48 per cent from light nuclei—see §9 and Table VII. It therefore seems certain that in the majority of the disintegrations of heavy nuclei the primary interaction is not confined to a pair of nucleons.

Any postulated mechanism of the primary interaction in which the rest-energy of the meson is regarded as being shared by one or two nucleons, the subsequent excitation resulting from secondary collisions, has to meet the following fundamental difficulty. The experimental results establish two salient facts : firstly, that the excitation energy is equal to a large fraction of the rest-energy of the meson ; and, second, that the proportion of energetic protons is small. From the point of view of an assumed primary interaction involving one or two nucleons, these two facts are contradictory ; for in order to keep the proportion of energetic secondary particles small it is necessary to assume a large mean free path for the primary nucleons in nuclear matter. They thus only infrequently make collisions in escaping from the nucleus, and will rarely produce projected protons. This assumption of a small probability of interaction has the second consequence, however, that the nucelus, being relatively transparent to the fast nucleons, will be given only a small average energy of excitation ; a conclusion which is in contradiction with the experimental results.

It also appears to be difficult to reconcile the present observations with the results of recent calculations by Tamor in which it is assumed that the primary interaction involves an α-particle " cluster " and leads to the production of a triton and a neutron. These calculations lead to the conclusion that 43 per cent of the nuclei should be excited with an energy of less than 40 MeV., in which case almost all this fraction would fail to produce a visible " star ". Because of the increased probability of the emission of neutrons at the beginning of the " evaporation ", the expected proportion may be put even higher—even 50 per cent or more. We have seen, however, that the experimental results give a figure of 36 per cent for the proportion of "starless" mesons in heavy nuclei; and the calculated " prong distribution ", using an excitation energy of 100 MeV., leads to a value of 21 per cent. Further, the neutron–triton model predicts a ratio of approximately 1 for the relative numbers of fast protons from light and heavy nuclei. On the other hand, experimental observations recorded in §9 give a value of 1:4 for this ratio.

The following alternative hypothesis appears to be in better accord with the experimental results :—Let us assume that, in the primary interaction, the rest-energy of the π^--meson is shared among three or four nucleons, charged and uncharged. The fast protons observed may then be attributed, at least in part, to protons which have taken part in this primary act. The initial distribution may take place, for example, in an α-particle group, and lead to its complete disruption. After the interaction, most of these four nucleons, of which one will be a proton, will be captured within the nucleus, for they are of relatively low energies, and the Ag and Br nuclei will be opaque to them. In particular circumstances, however, it will be possible for one of these nucleons to escape ; viz., if the following three conditions are satisfied after the interaction :—

(1) the momentum of the nucleon is directed outward ;
(2) the nucleon is sufficiently near the surface of the nucleus ; and
(3) the kinetic energy of the nucleon inside the nucleus is much greater than the maximum Fermi energy plus the binding energy of the nucleon.

Condition (3) may be written as

$$E_i > E_F + 8 + 30 (\text{in MeV.}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where E_i is the energy of internal motion and E_F the maximum Fermi energy, and the binding energy is taken as 8 MeV.; (30 MeV. is the minimum energy of the fast protons considered).

The Heisenberg relation then gives us (Rosenfeld 1948) that only those protons produced in a spherical shell of thickness $1{\cdot}6r_0$ and satisfying the above conditions will be observed as fast protons, r_0 being the "nucleon radius" $=1{\cdot}47\times10^{-13}$ cm. Thus we obtain for the number of fast protons with energy greater than 30 MeV.,

$$n(E_0>30)=n(E_i>E_F+8+30)\times\tfrac{1}{2}\times\frac{\text{volume of spherical shell}}{\text{volume of nuclear sphere}}, \quad . \quad (3)$$

E_0 being the energy of the proton observed outside the nucleus.

As a first approximation, the spectrum of E_i-E_F may be considered to be similar to that which would result from the sharing of the total energy of the π-meson between 4 nucleons of a free helion. The results of this interaction have been studied theoretically by the Liverpool group (Ruddelsdon and Clark 1949). Introducing into the right hand side of the equation (3) the proportion of particles in their calculated spectrum with energy greater than 38 MeV., it is found that 13 per cent of the π^--mesons absorbed in silver nuclei should be accompanied by the emission of fast protons of energy greater than 30 MeV. It will be possible to test the validity of the assumptions employed in obtaining this result when more experimental material becomes available, but the preliminary result appears to be in satisfactory agreement with the value for heavy nuclei obtained experimentally :—viz., 21 per cent.

An approximate value of the average energy of excitation may also be calculated on the basis of this model. The fraction of the energy carried away by the fast protons is about a quarter of the rest-energy of the meson, and the relative abundance of the fast protons is approximately 0·1 per star. The three neutrons have an equal probability of emission with the same energy, and therefore the amount of energy which is not used in "heating" the nucleus is equal to $0{\cdot}1\times4\times\frac{1}{4}\mu c^2$ where $\mu c^2=$rest mass energy of the meson.

This gives an average excitation energy of $\sim$120 MeV. after accounting for the binding energy of the fast nucleons emitted.

To a first approximation, the energy spectrum of the fast nucleons emitted, would be expected from this model to be similar to that resulting from the sharing of the rest-energy of the meson between 4 particles in the primary interaction, and to display a maximum at an energy between 30 and 40 MeV. A preliminary estimate of the ratio of the number of fast protons with energy greater than 50 MeV. to the number between 30 and 50 MeV., as deduced from the spectrum calculated by the Liverpool group, gives a value in good agreement with out experimental observations.

PART III.

Interaction of Slow π^--Mesons with Light Nuclei.

§16. GENERAL FEATURES OF THE DISINTEGRATIONS.

The capture of π^--mesons by light nuclei commonly leads to more complete disintegration than in the case of heavy nuclei. It therefore seemed probable that a detailed study of the characteristics of the corresponding stars would give important information complementary to that obtained from the observations described above.

In the conditions of our experiments, the stars representing the disintegration of light elements can only arise from the three types of nuclei ${}^{12}_{6}C$, ${}^{14}_{7}N$ and ${}^{16}_{8}O$. It is therefore possible to make a tentative identification of the disintegration of particular nuclei by an application of the principle of the conservation of charge. Although many 1- and 2-pronged stars occur in which the fragments are of charge three or more, and of which the parent nucleus is unknown, a sufficient number of stars with three or more prongs have been found to render this process useful. Among the latter, the following three types are most frequently observed:—

A, stars with 2 prongs of charge $|2e|$ and 1 prong of charge $|e|$ identified as due to the disintegration of ${}^{12}_{6}C$;

B, stars with 3 prongs of charge $|2e|$, identified as due to the disintegration of ${}^{14}_{7}N$; and,

C, stars with 3 prongs of charge $|2e|$ and 1 prong of charge $|e|$ identified as due to the disintegration of ${}^{16}_{8}O$.

The recognition of stars of these types is facilitated by the fact that in Ilford C2 emulsions, with normal processing conditions, it is usually possible to distinguish between singly and doubly charged particles by inspection if the track is longer than 10μ. It must, however, be borne in mind that the star types A, B and C are not the only modes of disintegration of the nuclei ${}^{12}_{6}C$, ${}^{14}_{7}N$ and ${}^{16}_{8}O$. Thus, 22 ± 7 per cent of the stars observed in the gelatine sandwiches (Table II.) are of the type A, yet on the basis of Fermi and Teller's formula, approximatley 43 per cent of the stars in the gelatine should be attributed to ${}^{12}_{6}C$ nuclei.

Using Fermi and Teller's theory, and taking account of the capture of π^--mesons by hydrogen nuclei, the proportion of the stars due to the various nuclei in the gelatine has been calculated. Since our results indicate that 45 ± 5 per cent of all the stars in the emulsion are formed in light nuclei, the fraction of all the stars which come from ${}^{12}_{6}C$, ${}^{14}_{7}N$ and ${}^{16}_{8}O$ nuclei may be determined. The results are shown in the second column of Table IX., whilst the fourth column shows the frequency of occurrence of stars of type A, B and C as observed in 780 σ-stars in C2 emulsions.

The analysis of the stars is complicated by the emission of neutrons which cannot be observed, but reactions A, B, C may be represented, tentatively, by the equations :—

$$\text{A,}\quad {}^{12}_{6}\text{C}+\pi^{-}\longrightarrow 2\,{}^{4}_{2}\text{He}+1\,{}^{1}_{1}\text{H}+3\,{}^{1}_{0}\text{n}\,;$$

$$\text{B,}\quad {}^{14}_{7}\text{N}+\pi^{-}\longrightarrow 3\,{}^{4}_{2}\text{He}+2\,{}^{1}_{0}\text{n}\,;$$

$$\text{C,}\quad {}^{16}_{8}\text{O}+\pi^{-}\longrightarrow 3\,{}^{4}_{2}\text{He}+1\,{}^{1}_{1}\text{H}+3\,{}^{1}_{0}\text{n}.$$

Among the various mechanisms which have been proposed to account for the disintegrations produced by the nuclear capture of a meson, the simplest is that the total available energy is symmetrically distributed amongst the products of the disintegration ; such a model was, we have seen, proposed tentatively, by the Liverpool group, for the case of the ${}^{4}_{2}\text{He}$ nucleus. The total energy is regarded as distributed among n particles in such a way that each has the same probability of receiving a certain fraction of the maximum possible energy ; a probability which is only proportional to the corresponding volume in configuration space. The magnitude of the energies involved is sufficiently great to allow the effect of binding forces to be neglected, except in considering the total balance of energy.

Table IX.

Calculated percentage * of the stars which originate in		Observed percentage * of the stars of type	
${}^{12}_{6}\text{C}$	20±2	A	7·6±1
${}^{14}_{7}\text{N}$	5·2±5	B	·3·7±7
${}^{16}_{8}\text{O}$	20±2	C	4·2±7

* Percentages are expressed in terms of the total number of stars in the emulsion.

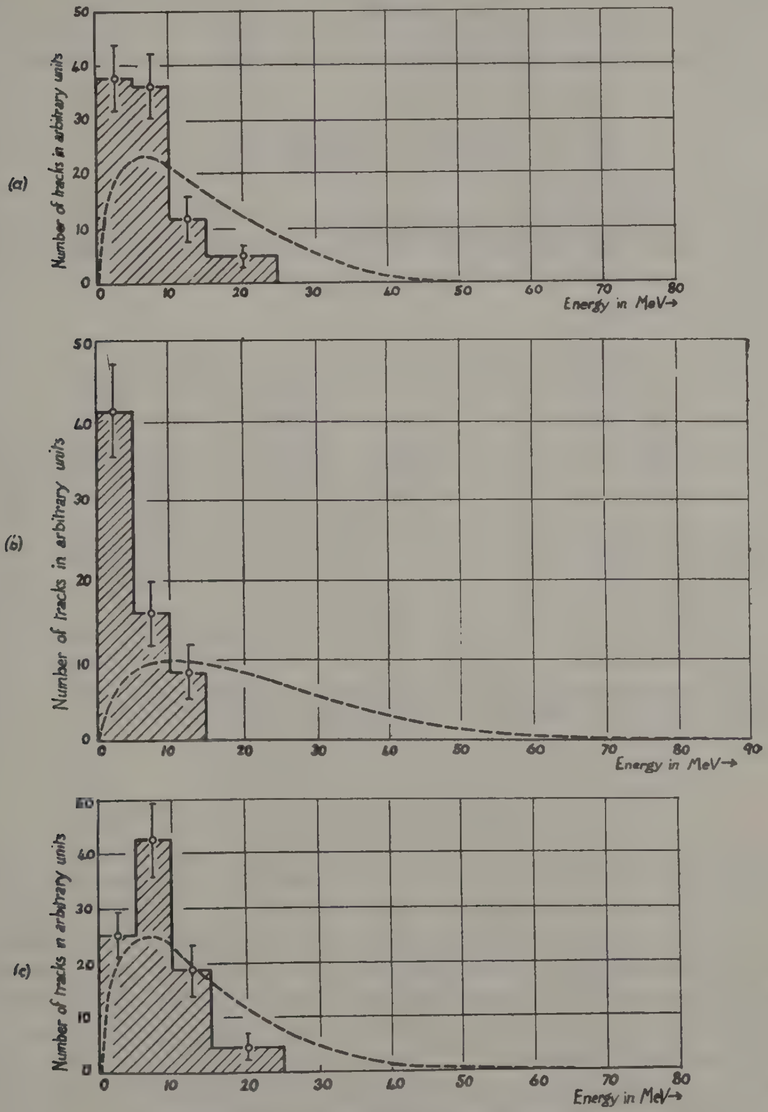
The above assumption leads to a distribution function of the form

$$\text{P(E)}\,d\text{E}=\text{E}_{\text{max}}^{\frac{5-3n}{2}}\,\frac{\Gamma\left(\frac{3n-3}{2}\right)}{\Gamma\left(\frac{3n-6}{2}\right)}\cdot\frac{2}{\sqrt{\pi}}\cdot\text{E}^{1/2}(\text{E}_{\text{max}}-\text{E})^{\frac{3n}{2}-4}\,d\text{E},\quad .\quad (4)$$

where n is the number of particles sharing the rest-energy of the meson ; and E_{max}, the maximum energy of the particle considered, equals $\text{E}_0(\text{A}-\text{M})/\text{A}$. In this expression, E_0 is the total available energy, A the mass of the nucelus, and M that of the particle considered.

Fig. 5 gives the distribution in energy of 102 α-particles form the stars of reaction A, 65 from stars of reaction B, and 100 from stars of reaction C. The dotted lines in the figure represent the distribution function (4)

Fig. 5.



Distribution in energy of α-particles.

(*a*) 102 α-particles from disintegration of $^{12}_{6}C$.

(*b*) 65 α-particles from disintegration of $^{14}_{7}N$.

(*c*) 100 α-particles from disintegrations of $^{16}_{8}O$.

Dotted curves give corresponding theoretical distributions using function (4) (see text).

calculated for $n=6$ in case A, $n=5$ in case B and $n=7$ in case C, values which correspond to the number of secondary particles in each of the reactions A, B and C if they are correctly represented by equations given above.

The disagreement between experimental and theoretical distributions in fig. 5 indicates that there is a large disparity in the distribution of the rest-energy of the meson among the ejected particles. The same conclusion is reached by considering their average energy which is given by equation

$$\int_0^{E_{\max}} E \, . \, P(E) \, dE = \bar{E} = [1/(n-1)] E_{\max} . \qquad (5)$$

Table X. shows the calculated and observed values of the energy of these particles, and it will be seen that the theoretical values of the energy of the neutrons are much lower than those deduced from the experimental results.

Table X.

Reaction	$\bar{E}_\alpha$ observed (MeV.)	$\bar{E}_\alpha$ calculated (MeV.)	$\bar{E}_p$ observed (MeV.)	$\bar{E}_p$ calculated (MeV.)	$\bar{E}_n$ * observed (MeV.)	$\bar{E}_n$ † calculated (MeV.)
A	5·8	13·6	8·6	18·7	~82/3	56/3
B	5·0	21	—	—	~105/2	56/3
C	7·5	11·8	~8	14·8	64/3	44·4/3

* Difference between the total available energy and the average total energy in the charged particles divided by the number of neutrons.

† The average calculated energy of one neutron.

§ 17. Capture by $^{14}_{7}$N.

The second row of Table X. shows that in the disintegration of $^{14}_{7}$N, most of the energy is imparted to two neutrons, and that the three α-particles receive a total kinetic energy of only approximately 15 MeV. In this case, an attempt has been made to make a closer analysis of the reaction by considering the momentum balance, although the nature of the charged particles, assumed to be α-particles, is not always certain.

If the three α-particles represent the visible products of the disintegration of $^{14}_{7}$N, the resultant of their momenta gives some information about the energies and directions of ejection of the two neutrons. The The momenta have been determined for those cases in which all three α-particles remain in the emulsion. In such cases, it is possible to determine upper and lower limits for the energy of the two neutrons, limits which correspond to the maximum and minimum possible values of the angle between their directions of ejection. The results are shown in Table XI.

Let us choose to consider, first, those limiting cases in which there is the greatest disparity between the energy of the two neutrons. Most of the rest-energy of the π-meson must then be assumed to have been imparted to one neutron, whilst the other may be considered as resulting from the disintegration of the residual nucleus. The characteristics of the observed disintegrations are such, however, that this slow neutron must be assumed to have been emitted in a direction making a small angle to the vector representing the sum of the momenta of the three α-particles. This result implies either that the neutron has a very small velocity of ejection compared to the velocity of the recoiling nucelus, or that it has a preferential direction of emission with respect to the line of motion of the nucleus. Both features are inconsistent with the view that the neutron arises in the spontaneous break up of a moving excited nucleus.

TABLE XI.

Total energies of the α-particles in MeV.	Maximum and minimum values for the angle of separation between the two neutrons	Limits of energy of each of the two neutrons	
		First	Second
15·4	$154° < \theta < 180°$	$51{\cdot}2 < E_1 < 74{\cdot}8$	$27{\cdot}6 < E_2 < 51{\cdot}2$
16·5	$130° < \theta < 180°$	$50{\cdot}8 < E_1 < 89{\cdot}3$	$12{\cdot}3 < E_2 < 50{\cdot}8$
17·3	$130° < \theta < 180°$	$52{\cdot}4 < E_1 < 92{\cdot}9$	$11{\cdot}9 < E_2 < 52{\cdot}4$
14·3	$132° < \theta < 180°$	$51{\cdot}9 < E_1 < 90{\cdot}3$	$13{\cdot}5 < E_2 < 51{\cdot}9$
20·7	$155° < \theta < 180°$	$48{\cdot}6 < E_1 < 69{\cdot}2$	$28 < E_2 < 48{\cdot}6$
10·6	$94° < \theta < 180°$	$53{\cdot}7 < E_1 < 107$	$0{\cdot}7 < E_2 < 53{\cdot}7$
9·0	$142° < \theta < 180°$	$54{\cdot}5 < E_1 < 91$	$17{\cdot}8 < E_2 < 54{\cdot}5$
7·3	$167° < \theta < 180°$	$55{\cdot}3 < E_1 < 65{\cdot}7$	$45{\cdot}9 < E_2 < 55$

In the second extreme case the two neutrons are approximately equal in energy. We must then visualize the production of two fast neutrons moving in nearly opposite directions and carrying most of the rest energy of the π^--particle. The remaining part of the disintegration may then be described in terms of an excited $^{12}_{6}C$ nucleus disintegrating whilst in motion. With this assumption, the excitation energies of the carbon nucleus can be calculated, and the values obtained lie between 16 and 25 MeV. These values are in good agreement with the results of Hanni, Telegdi and Zunti (1948), who have shown that the formation of three α-particles is the most common mode of disintegration of $^{12}_{6}C$ nuclei, excited to energies in the region of 20 MeV.

§18. CAPTURE BY ${}^{12}_{6}$C.

Because of the α-particle structure of the ${}^{12}_{6}$C nucleus, the possibility exists that, following the capture of a meson, the primary distribution of energy may take place among nucleons of one of the α-particles, leaving the rest of the nucleus unaffected. If the result of this interaction is a triton and a neutron (Tamor 1950), stars corresponding to the reaction

$$ {}^{12}_{6}\mathrm{C}+\pi^{-}\rightarrow {}^{4}_{2}\mathrm{H}+{}^{4}_{2}\mathrm{He}+{}^{3}_{1}\mathrm{H}+{}^{1}_{0}n \quad . \; . \; . \; . \; . \; . \; . \quad (6) $$

could occur, either through the escape of the "primary" triton with energy 30 MeV. or as a result of the escape of the "primary" neutron of 90 MeV. and the formation of a residual ${}^{11}_{5}\mathrm{B}^{*}$ nucleus excited to about 30 MeV. The latter nucleus is known to disintegrate into two α-particles and a triton at high excitation energies (Lattes and Occhialini 1947, Goward, Titterton and Wilkins 1950). For such values of the excitation energy, the concept of a compound nucleus is, of course, very crude, and the mode of decay probably depends on the mechanism of excitation. Although the study of the disintegrations of heavy nuclei, discussed in a previous section, indicates that an interaction giving rise to a triton and a neutron occurs only infrequently, a disintegration of carbon giving two α-particles and a triton would, although occurring rarely, be of considerable interest because the only particle escaping detection is a neutron; and it is therefore possible, in principle, to make a complete investigation of the energy and momentum balance.

The possibility of identifying such a reaction is complicated by the difficulty of establishing whether ejected particles of charge $|e|$ are protons or heavier hydrogen nuclei; but if a track is of sufficient length, identification is possible by measuring its grain-density, and the scattering of the particle. If, further, a particular disintegration is correctly represented by equation (6), there must be a consistency between the values of the energy and momenta of the unobserved neutron as determined by two independent methods :—(*a*) from an application of the principle of the conservation of energy, the energy of the ejected charged particles being known; and (*b*) by determining the resultant of the momenta of the three charged particles, and, hence, the momentum and energy of the neutron.

In view of this possibility, a complete analysis was made of 13 σ-disintegrations, of the type under consideration in which all the tracks end in the emulsion, and details are given in Table XII. In two cases, the mass of the particle of charge $|e|$, as determined by observations on the grain density along the track, appeared to be greater than that of a proton. Making the tentative assumption that all the particles of charge $|e|$ are tritons, values of their energy and momentum have been determined from their observed range. The corresponding values of the total energy of the particles in each star, and the vector sum of their momenta are given in cols. 3 and 4, respectively of Table XII., while col. 5 gives the momentum of the neutron assumed to carry the remaining available energy. The limits of error indicated include those due to straggling, to lack of precise knowledge regarding the shrinkage factor of the emulsion, and to errors in the measurement of the lengths and directions of the tracks. All errors have been considered for the most unfavourable cases, and are therefore much greater than the probable errors.

TABLE XII.

Energies of particles in the star	Evidence about nature of the particle of charge $\lvert e \rvert$	Total kinetic † energy of three charged particles (MeV.)	Total momentum † of the three charged particles ($\sqrt{2mE}$)	Calculated † momentum of the neutron $= \sqrt{2(140 - Q - \Sigma E_{kin})}$
α_1: 3·3 MeV. α_2: 6·5 MeV. p: 83 μ	?	14·5±1·3	4·6±0·9	13·8±0·2
α_1: 4·4 MeV. α_2: 2·3 MeV. p: 168 μ	?	14·2±1·3	6·5±0·4	13·8±0·2
α_1: 1·8 MeV. α_2: 1·7 MeV. p: 1232 μ	Identified as proton by scattering and grain density	27·5±1·7	5·4±2·9	12·7±0·3
α_1: 9·4 MeV. α_2: 7·0 MeV. p: 220 μ	?	24·4±1·0	7·9±3·5	13·1±0·1
α_1: 5·6 MeV. α_2: 5·6 MeV. p: 165 μ	?	20·7±1·5	10·2±1·4	13·5±0·2
α_1: 4·2 MeV. α_2: 3·5 MeV. p: 66 μ	?	12·0±1·3	7·1±1·1	14·0±0·2
α_1: 11·6 MeV. α_2: 14·0 MeV. p: 330 μ	Identified as proton	38·7±1·1	24±12	12·0±0·02
α_1: 2·4 MeV. α_2: 11·6 MeV. p: 945 μ	Identified as proton	35±1·4	16·9±5·0	12·2±0·2
α_1: 3·5 MeV. α_2: 0·5 MeV. p: 1425 μ	Identified as proton	31·3±1·6	17·7±1·8	12·5±0·2
α_1: 5·1 MeV. α_2: 5·8 MeV. p: 110 μ	* ?	16·5±1·3	14·3±1·0	13·7±0·2
α_1: 2·0 MeV. α_2: 5·8 MeV. p: 315 μ	* Identified as proton	18·8±1·3	14·7±1·0	13·5±0·1
α_1: 3·4 MeV. α_2: 9·0 MeV. p: 116 μ	* ?	18·3±1·0	12·4±6·0	13·1±0·2
α_1: 4·0 MeV. α_2: 7·9 MeV. p: 65 μ	* Grain density indicates triton possible	15·6±1·2	13·5±2·5	13·7±0·2
α_1: 13·0 MeV. α_2: 10·5 MeV. p: 74 μ	* Grain density indicates triton possible	28·2±1·0	14±6	12·8±0·2

† These columns have been computed, assuming that the particle of charge $\lvert e \rvert$ is a triton.

* In these cases, energy and momentum balance considerations seem to justify the assumption that the particle of charge $\lvert e \rvert$ is a triton.

From these observations, it may be concluded, first, that if tritons of energy 30 MeV. are somtimes produced as a result of the capture of π^--particles by carbon nuclei, they are rare. Second, in five of the disintegrations referred to in Table XII., there is a consistency between the balance of energy and of momentum, if the particle of charge $|e|$ is assumed to be a triton. In one of these examples, however, the particle of charge $|e|$ can be identified as a proton so that the apparent consistency between the momentum and energy balance in the star must be regarded as fortuitous. In two of the remaining four examples, the grain-densities in the tracks are consistent with the view that the particles of charge $|e|$ are more massive than protons. For these four cases the average energy of the neutron required to provide the balance of energy and momentum is 85 MeV., and the neutron–triton model appears to give a satisfactory description of the disintegrations.

The above results suggest that in a small proportion of the disintegrations of carbon nuclei, the transmutation can be correctly represented by equation (6). Many of the other examples of which the details are included in Table XII. may be attributed to an interaction with an α-cluster in which its disintegration is more complete than that suggested by the neutron–triton model. In a few cases, however, the following analyses suggest that the primary interaction is not confined to a single α-cluster.

The energy of the particles observed in the $2a+1p$ stars has been plotted in a triangular energy diagram ; see fig. 6. In this diagram the three coordinates are (*a*) the total energy of the two α-particles ; (*b*) the energy of the particle of charge $|e|$, assumed to be a proton ; and (*c*) the remaining available energy which represents the energy carried by the three neutrons. The sum of the three coordinates of any point in this diagram is 100 MeV. This value has been chosen because it is nearly equal to the difference between the rest-energy of the π-meson and Q value of the reaction ${}^{12}_{6}\mathrm{C} \rightarrow 2\alpha+1p+3n$.

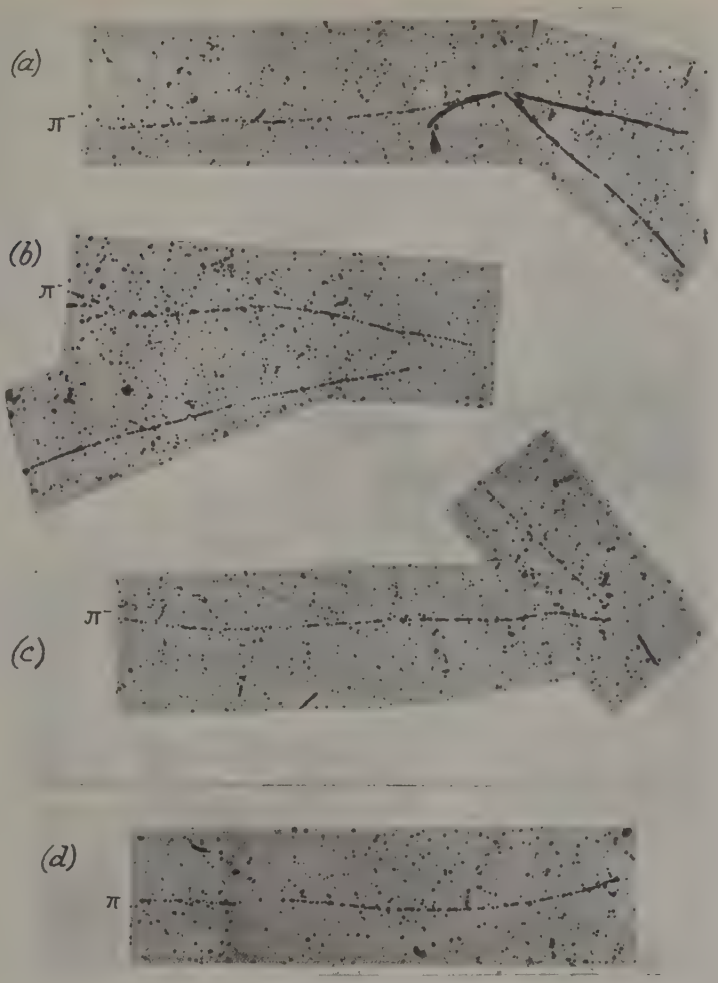
It will be seen that the points in fig. 6 are grouped together in a region corresponding to small values of the α-particle and proton energies, but that there are some points widely displaced from it. These correspond to disintegrations in which the α-particles carry away most of the rest-energy of the meson. The effect is particularly marked for the two examples described below ; the letter α_1 and α_2 refer to the two α-particles, and p to the proton.

Case 1 : α_1 24 MeV., stays in emulsion;
α_2 : 19 MeV., stays in emulsion;
angle between α_1 and α_2 153°;
p : leaves the emulsion after traversing a path length of 100μ.

*Case 2 : α_1 ; 21·5 MeV., stays in emulsion;

---

* This star was found recently by Dr. D. H. Perkins, who kindly directed our attention to it.



CAPTIONS FOR PLATES.

Plate XVIII.

(*a*), (*b*), (*c*). Three typical examples of disintegrations observed in the "sandwich-plates". A meson is captured by a light element and produces a star whose centre is absent because it is located in the layer of pure gelatine 5 μ thick.

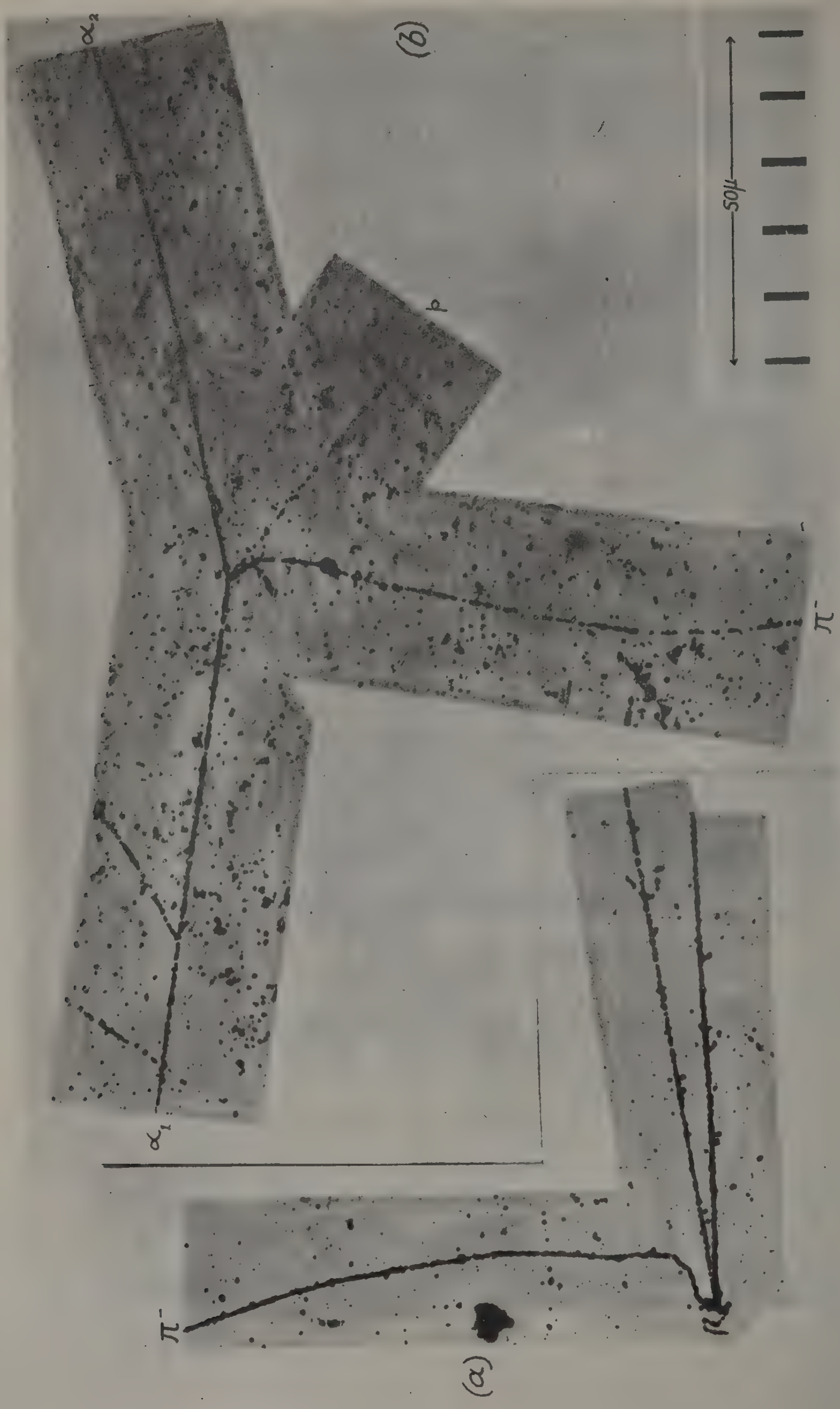
(*d*) Meson showing a "blob". Events of this type are sometimes observed when negative mesons come to the end of their range in the emulsion and do not produce a star.

Plate XIX.

(*a*) Sigma-star accompanied by a slow electron. (Auger effect in heavy nuclei.)

(*b*) Energetic "$2\alpha+1p$" star identified as the disintegration of a ^{12}C nucleus. Two alpha-particles of 24 and 19 MeV. respectively, are emitted at an angle of 153° to each other.

This star represents a possible case in which the capture process leads directly to the total disruption of a light nucleus.

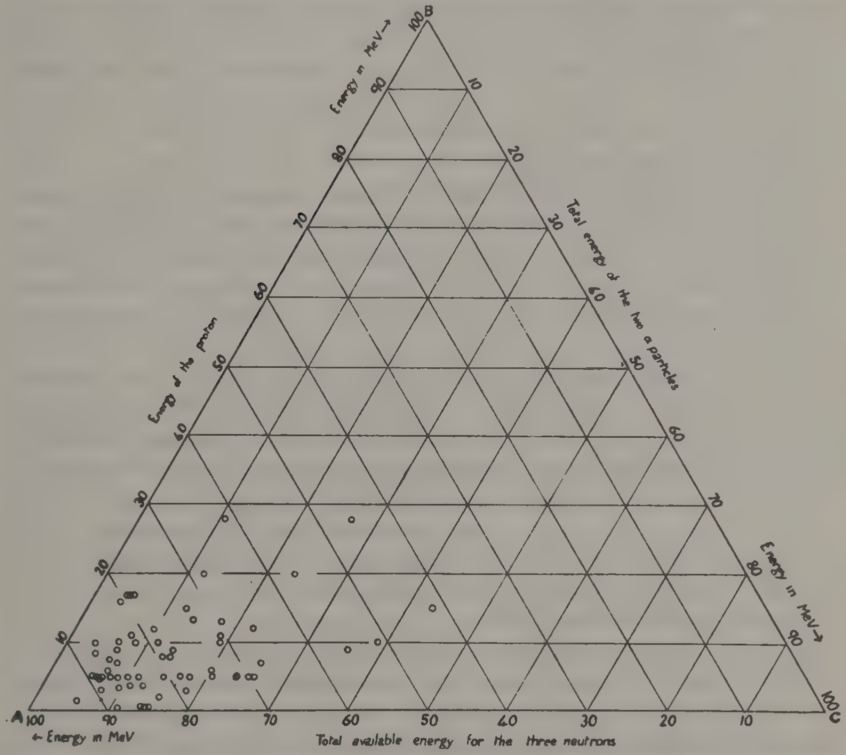

(b)
α_2
p
50μ
π^-
α_1
π^-
(a)

α_2 : 5·4 MeV., stays in emulsion;

p : leaves the emulsion after a path length of 3250μ ; energy from grain-count and multiple scattering 28±3 MeV.; angle between α_1 and $p=180°$.

A photo-micrograph of case 1 is shown in Pl. XIX.

Fig. 6.



Distribution of the available energy amongst the disintegration fragments resulting from the disintegration of $^{12}_{6}C$.

The total energy of the two α-particles is constant along lines parallel to AB. Similarly, the energy of the proton and total available energy for the 3 neutrons is constant along lines parallel to AC and CB respectively.

These disintegrations appear to indicate that a larger part of the nucleus can sometimes take part in the primary interaction, and that the α-particles can interact as a whole group in the distribution of energy. It would seem, therefore, that the distance between the α-particle groups in a carbon nucleus is sufficiently small to permit this interaction to compete with the disintegration of a single α-cluster.

§19. COMPARISON OF THE DISINTEGRATIONS OF LIGHT AND HEAVY NUCLEI.

The present observations give support to the conclusion reached by previous observers that "radiative" capture, or the emission of neutrettos, rarely or never accompanies the capture of π^--particles by nuclei of charge number greater than six. It follows that several nucleons are involved in the primary act and one can, as a rough approximation, assume that all the nucleons situated within a certain distance of the interacting proton will be involved in the interaction. This distance may be computed, for instance, by applying the uncertainty relation (Bruno 1949)

$$\Delta r = c\Delta\tau \quad \text{with} \quad \Delta\tau \times \Delta E = \hbar \quad \text{and} \quad \Delta E = 140 \text{ MeV.};$$

the result is $r = 1{\cdot}35\ 10^{-13}$ cm.

There is, therefore, on this view, a certain volume of nuclear matter influenced by the primary interaction, and the dimensions of this volume are more or less constant and independent of the size of the nucleus. The difference between the characteristics of the disintegrations produced in light and heavy nuclei can then be explained qualitatively in the following terms :—

In heavy nuclei the primary interaction is confined to a relatively small part of the nucleus, and most of the fast nucleons originating from it lose their energy in the surrounding mass of the nucleus. When the interaction takes place near the edge of the nucleus, however, a nucleon may sometimes escape, either directly or after making one collision with a nucleon. Apart from these fast particles, the disintegration of the nucleus is independent of the precise characteristics of the primary process and takes place many nuclear periods afterwards, as a result of the rise in "temperature" of the nucleus.

In light nuclei the primary process involves a much larger part of the nucleus. The observed disintegration is likely to be much more dependent upon particular characteristics of the primary process, and especially on the number and the nature of the fragments which take part in the initial sharing of energy. This explains the occurrence of stars, originating from nuclei of the same type, which show wide variations in the total energy represented by the ejected charged particles. Owing to their strong cohesion the α-groups are likely to participate as a whole in the interaction, a fact which may explain the low probability of observing fast isolated protons from the disintegration of light nuclei.

CONCLUSIONS.

(*a*) The separation of σ-stars produced in light (C, N, O) and heavy (Ag, Br) nuclei in photographic emulsions has been carried out using three different methods. The results indicate that 46 ± 5 per cent of all visible σ-stars are due to the interaction of the π^--mesons with light nuclei, and 54 ± 5 per cent with heavy nuclei. 28 ± 2 per cent of all the π^--mesons absorbed suffer capture without producing any visible star.

(*b*) The relative frequency with which the disintegrations are accompanied by the emission of different numbers of charged particles has been determined separately for light and heavy nuclei. For light nuclei there is a maximum in the distribution at a point corresponding to the emission of three charged particles. For heavy nuclei, on the other hand, there is a rapid fall as the number of emitted particles increases, disintegrations with no charged particles emitted occurring most frequently.

(*c*) Nuclear reactions produced by the capture of π^--mesons in heavy nuclei exhibit characteristics most of which are in good accord with the general predictions of the Weisskopf evaporation theory. The average number of prongs observed per heavy nucleus is $1{\cdot}1 \pm {\cdot}15$. The α/p ratio, the ratio of the number of emitted α-particles to the number of protons, is $0{\cdot}30 \pm 0{\cdot}04$. The absence of any appreciable motion of the centre-of-mass system commonly prevents the observation of the recoil nuclei. The energy spectrum of the ejected protons has the characteristic Maxwellian form for particles of energy up to 20 MeV., and is consistent with an average excitation energy of 109 MeV. This value is in good accord with the observed prong distribution and α/p ratio.

The energy spectrum of the ejected protons departs from a Maxwellian distribution at high energies. $21{\cdot}4 \pm 2{\cdot}5$ per cent of the σ-stars in heavy nuclei have an associated fast particle of charge $|e|$ with an energy greater than 30 MeV. An interpretation is offered according to which these protons represent, "primary" protons resulting from the distribution of the rest-energy of the π^--mesons among three or more nucleons, or charged secondaries produced in the collisions with protons in the nucleus of fast "primary" neutrons.

(*d*) By applying the principle of the conservation of charge it has been found possible to identify the parent nucleus in about half the stars due to light elements; and to study in particular the disintegration of ^{12}C and ^{14}N.

The nuclear reactions produced by the capture of π^--mesons in light elements show the following characteristics :—

(1) The rest-energy of the meson is not dispersed symmetrically amongst the products of the disintegration. In the majority of cases the primary interaction results in one or two fast neutrons and an excited residual nucleus.

(2) In the case of the disintegration of the $^{14}_{7}$N nucleus the states of excitation of the residual $^{12}_{6}$C nucleus into 3α particles have been measured.

(3) The presence in a few stars of high energy α-particles with energies exceeding 20 MeV. indicate that these particles may occasionally take part as a whole in the initial distribution of the rest-energy of the mesons.

(4) No tritons with energy $\sim$30 MeV., such as would result from the interaction of the π^--meson with an α-particle group on the neutron–triton model, have been observed. The occurrence of slow tritons from such a process, though unlikely, is not excluded.

(5) Particles of charge $|e|$ with energies greater than 30 MeV. have been observed in $4{\cdot}9 \pm 1{\cdot}5$ per cent of all the σ-stars produced in light nuclei.

ACKNOWLEDGMENTS.

We have pleasure in thanking Professor C. F. Powell, F.R.S., for his continued interest and encouragement during the course of the work. We are especially grateful to Dr. E. Gardner and Professor C. M. G. Lattes for arranging exposures of our plates in the Berkeley cyclotron, to Dr. H. Heitler for technical assistance and Mr. C. Waller of Ilford Ltd., for supplying us with the sandwich plates used in this investigation. We should also like to thank Professor H. Fröhlich and his group at the University of Liverpool and Mr. U. Camerini of this Laboratory, for helpful discussions.

One of us (O.R.) would like to take this opportunity to express his gratitude to the British Council for the award of a scholarship which has enabled him to pursue research at Bristol.

REFERENCES.

ADELMAN and JONES, 1949, *Phys. Rev.*, **75,** 1468 (A).
BRUECKNER and POWELL, 1949, *Phys. Rev.*, **75,** 1274.
BRUNO, 1948, *Ark. f. Mat. Astro. Fys.*, A. **36,** 1 ; 1949, *Ark. f. Fys.*, **1,** 19.
CHESTON and GOLDFARB, 1950, *Bull. Amer. Phys. Soc.*, **25** (1) (Abstract D 5).
CLEMENTEL and PUPPI, 1949, *Nuovo Cim.*, **6,** 1.
COSYNS, DILWORTH, OCCHIALINI, SCHOENBERG and PAGE, 1949, *Proc. Phys. Soc.* A, **62,** 801.
FERMI and TELLER, 1947, *Phys. Rev.*, **72,** 399.
FOWLER, 1950, *Phil. Mag.*, **41,** 169.
FRANZINETTI, 1950, *Phil. Mag.*, **41,** 86.
FUJIMOTO, HAYAKAWA and YAMAGUCHI, 1949, *Prog. Theor. Phys.*, **4,** 576.
FUJIMOTO and YAMAGUCHI, 1949, *Prog. Theor. Phys.*, **4,** 468.
GOLDBERGER, 1948, *Phys. Rev.*, **74,** 1269.
GOWARD, TITTERTON and WILKINS, 1950, *Proc. Phys. Soc.* A, **63,** 172.
HANNI, TELEGDI and ZUNTI, 1948, *Helv. Phys. Acta.*, **21,** 203.
HARDING, 1949, *Phil. Mag.*, **40,** 530.
HEIDMANN and LEPRINCE-RINGUET, 1948, *Comptes Rendus*, **226,** 1716.
HEISENBERG, 1937, *Leipziger Ber.*, **89,** 369.
LATTES and OCCHIALINI, 1947, *Nature, Lond.*, **159,** 331.
LE COUTEUR, 1950, *Proc. Phys. Soc.* A, **63,** 259.
MARSHAK, 1949, *Phys. Rev.*, **76,** 114.
MARSHAK and WHITMAN, 1950, *Bull. Amer. Phys. Soc.*, **25** (3) (Abstract M 6).
PANOFSKY, 1950, New York Meeting of the Amer. Phys. Soc.
PERKINS, 1948, *Nature, Lond.*, **161,** 487 ; 1949, *Phil. Mag.*, **40,** 601 ; 1950, *Ibid.*, **41,** 138.
OCCHIALINI and POWELL, 1948, *Nature, Lond.*, **162,** 168.
ROSENFELD, 1948, *Nuclear Forces* (Amsterdam : North Holland Publishing Co.), pp. 266.
RUDDLESDON and CLARK, 1949, *Nature, Lond.*, **164,** 487.
TAMOR, 1950, *Phys. Rev.*, **77,** 412.

LV. *Notices of New Books and Periodicals received.*

The Nature of Physical Theory. By P. W. Bridgman. (Dover Publications, Inc.). 1949.

The presentation of physical theory now generally accepted errs, according to Professor Bridgman, because interest centres on the mathematical equations summarizing physical knowledge, whilst inadequate attention is paid to the context within which these equations take on their significance—the context of the methodology of experimental physics. The "operational" viewpoint, first expounded in the author's earlier 'The Logic of Modern Physics', is extended in 'The Nature of Physical Theory' to more metaphysical, and therefore more controversial aspects of the subject, only briefly touched upon in his earlier work. It is to be hoped that, when discerning readers disagree with some of the arguments of the book, as is inevitable in so subtle and neglected a subject, they will thereby be stimulated to further thought upon it. Unquestionably, Professor Bridgman's critical examination of the principal physical theories makes it evident that they possess disquieting features of which most physicists will be surprised to learn. The book has a particularly important bearing on the validity of the rapidly developing subject of cosmology. W. R. H.

Electronic Engineering Master Index, 1947–48. Editor, John F. Rider. Price $19.50.

This is a most useful bibliography covering publications in electronics and allied subjects in the two years 1947 and 1948. It includes also patent specifications, and gives titles, authors and sources of the articles listed, but no abstracts. There is a cross index which refers also to previous editions. It might be criticized in its attempt to cover some subjects—for example, the entries headed "Nucleonics" comprise only about a page and seem to be chosen in an entirely arbitrary manner, but in the field of electrical communication and the more closely allied subjects, the entries appear to be very well classified and easy to locate. D. F. G.

Numerical Calculus. By W. E. Milne. [Pp. x+393; $6\frac{1}{4}'' \times 9\frac{1}{4}''$.] (Princeton University Press; London: Geoffrey Cumberlege, 1949.) English price 21*s*. net.

This book aims to bridge the gap between formal pure mathematics and detailed numerical answers. It covers the solution of simultaneous equations, approximation, interpolation, finite differences, numerical integration and differentiation, solution of differential equations, and curve-fitting by least squares. The book is essentially practical, and numerous examples are worked by methods best adapted to machine computation. While essential theory is given, proofs are often omitted or only briefly indicated ; the method for finding complex roots could have been explained in more detail.

The book is clearly written and is thoroughly recommended both as a text for undergraduate courses in numerical methods and for scientists who are forced to seek numerical solutions. The casual user will find the classified guide to formulæ and methods in appendix C particularly useful when searching for a formula or method suited to a particular problem.

The text has been photographically reproduced from typescript and contains a number of trivial errors which could have been removed before printing. J. K. M.

Photoelasticity. By H. T. Jessop and F. C. Harris. (Cleaver-Hume Press Ltd.)

This is a good text-book for those wishing to learn the potentialities of the photoelastic method of stress analysis, as well as for those engaged in photoelastic work. It answers many of the questions which are often asked by those beginning to study the subject, and the principles are clearly explained without recourse to difficult mathematics. The book contains much helpful guidance in experimental methods, the fruits of many years' experience. As an example, the chapters on elementary optics and on the optical systems used in photoelasticity should be valuable to anyone setting up a photoelastic bench or wishing to improve the efficiency of an existing one. Consideration is given to the comparatively recent "frozen stress" method of three-dimensional stress analysis, of which the usefulness and limitations are discussed. A number of examples of the application of photoelastic methods to specific problems are given and the book includes tables of the mechanical and photoelastic properties of the materials used. The addition of a bibliography would have been useful to those wishing to make a more detailed study of the subject. C. A.

Tables of Inverse Hyperbolic Functions. By The Staff of the Computation Laboratory. [Pp. xx+290; $8\frac{1}{4}'' \times 10\frac{3}{4}''$.] (Harvard University Press; London: Geoffrey Cumberlege, 1949.). English price 55*s.* net.

The tables form volume 20 of "The Annals of the Computation Laboratory of Harvard University", and required a mere twelve days for computing and printing by the Automatic Sequence Controlled Calculator. They contain tables of the inverse hyperbolic sine, cosine, and tangent to nine decimal places, together with first and second forward differences; the interval in each part of the tables is chosen so that second-order interpolation gives the full nine decimal accuracy over almost the whole of the tables.

There are four tables: Table I. gives $\tanh^{-1} x$ for $x=0(0{\cdot}001-0{\cdot}00001)0{\cdot}99999$; Table II. gives $\sinh^{-1} x$ for $x=0(0{\cdot}002,\ 0{\cdot}005)3{\cdot}495$; Table III. gives $\cosh^{-1} x$ for $x=1{\cdot}00000(0{\cdot}00001-0{\cdot}002)3{\cdot}498$; Table IV. gives $\sinh^{-1} x$ and $\cosh^{-1} x$ for $x=3{\cdot}5(0{\cdot}005-20)22{,}980$, while for larger values of x these two functions differ from $\log_e 2x$ by less than the round-off error in the tables. The error in the tables is estimated to be less than 0·6 unit in the ninth decimal in Table I. and less than 0·9 unit in the ninth decimal in the other tables.

These tables are quite unique and should be of great use to physicists, statisticians, engineers and applied mathematicians generally. J. K. M.

The Variational Principles of Mechanics. By Cornelius Lanczos. (Toronto University Press. London: Geoffery Cumberleye.) [Pp. 307.] Price 42*s.*

It is stated in the preface that "the present treatise on the variational principles of mechanics should not be regarded as competing with the standard text-books on advanced mechanics . . . the author feels that there is room for monographs which exhibit the fundamental skeletons of the exact sciences in an elementary and philosophically orientated fashion". This statement is probably the simplest summary of a work which, starting from the principle of virtual work and D'Alembert's principle, derives the variational principles of Gauss, Hamilton, Jacobi and Least Action, and goes on to determine Lagrange's equations, and the canonical equations of motion and to discuss canonical transformations and the Hamilton–Jacobi equation. Although the present reviewer's view-point differs in some fundamental points from the author's, he found that the author's obvious enthusiasm had made the work eminently readable and likely to provide the student of the subject with a useful exposition. L. H.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LVI. *The Thermal Conductivity of Potassium Chrome Alum at Temperatures below One Degree Absolute.*

By C. G. B. Garrett,
Royal Society Mond Laboratory, Cambridge *.

[Received April 12, 1950.]

Abstract.

An account is given of measurements of the thermal conductivity of potassium chrome alum made between 0·14° and 0·30° K., from which it is concluded that the mean free path of the phonons in the crystal is about ½ mm. In the method employed, a long single crystal of the salt, isolated from the outside world, is demagnetized to a uniform low temperature: a temperature gradient along the crystal is then set up, and the time constant of the subsequent heat flow measured. The paper concludes with a consideration of the lowest temperature to which a sample of material could in a reasonable time be cooled by direct thermal contact with a cold mass of paramagnetic salt.

§ 1. Introduction.

Temperatures below 1° K. may be obtained by the process of magnetic cooling, in which a paramagnetic salt is used as the refrigerant. In order to extend the study of these very low temperatures to other materials it is desirable to have information about the time required for cooling by direct contact with the cooled salt, and one of the controlling factors in this process will be the thermal conductivity of the paramagnetic crystals themselves. In the following paper an account is given of measurements of the thermal conductivity of a typical paramagnetic salt, potassium chrome alum, in the range of temperature between 0·14° and 0·30° K. Such measurements are also of interest in connection with the theory of heat conduction in insulating materials.

Precise measurements of the thermal conductivity of potassium chrome alum have been made in the liquid helium temperature range by Bijl (1949), but below 1° only the order of magnitude of the conductivity is known (Kürti, Rollin and Simon 1936). The latter determination was made by demagnetizing a long single crystal of the salt from an inhomogeneous field, so that the ends were cooled to somewhat different

---

* Communicated by Sir Lawrence Bragg, F.R.S.

temperatures: the subsequent approach to temperature equality was watched by means of local magnetic susceptibility measurements made with mutual inductances placed around the two ends, and the conductivity estimated from semi-qualitative considerations. This method has the intrinsic advantage that it does not require separate thermometers or a heating coil attached to the crystal, thereby avoiding trouble from contact thermal resistances and heat inflow along metallic leads. In the present investigation a similar method has been used, in which however a device is employed for setting up a temperature gradient whenever required, without the necessity for a fresh demagnetization. In addition a more exact analysis of the heat flow has been made, and care has been taken to keep temperature differences as small as possible, with a view to investigating the temperature variation of the conductivity over the observable range.

§ 2. PRINCIPLE OF THE METHOD.

The salt, in the form of a long single crystal, is demagnetized from a uniform magnetic field, so that its final temperature is uniform. A non-uniform magnetic field, obtained from a suitable system of coils, is applied to the crystal in such a way that the field varies linearly along the length of the crystal with a gradient of the order of 100 gauss/cm. A temperature gradient is set up and heat flows from the warmer to the cooler end of the crystal. After allowing time sufficient for the temperature to become once more approximately uniform, the non-uniform field is removed, and a temperature gradient is now set up in the opposite sense to the first. The thermodynamics of this process will be discussed in § 3. The temperatures near the two ends of the crystal may be measured by systems of mutual inductance coils suitably placed, and the rate of approach of those temperatures determined*.

From the rate of approach of temperature it is possible to calculate the diffusivity of the material, $K/\rho C$, where K is the thermal conductivity, ρ the density and C the specific heat per unit mass. The value of C appropriate to the mean temperature of the crystal may be taken from previous measurements, and hence the thermal conductivity may be derived. Since, in the temperature range above 0·2°, $C \sim A/T^2$, and since the conductivity itself is proportional approximately to T^3, the diffusivity, and hence the equilibrium time of the flow process, are extremely rapidly varying functions of temperature. It is therefore possible, for a given size of crystal sample, to make measurements only over a limited temperature range.

---

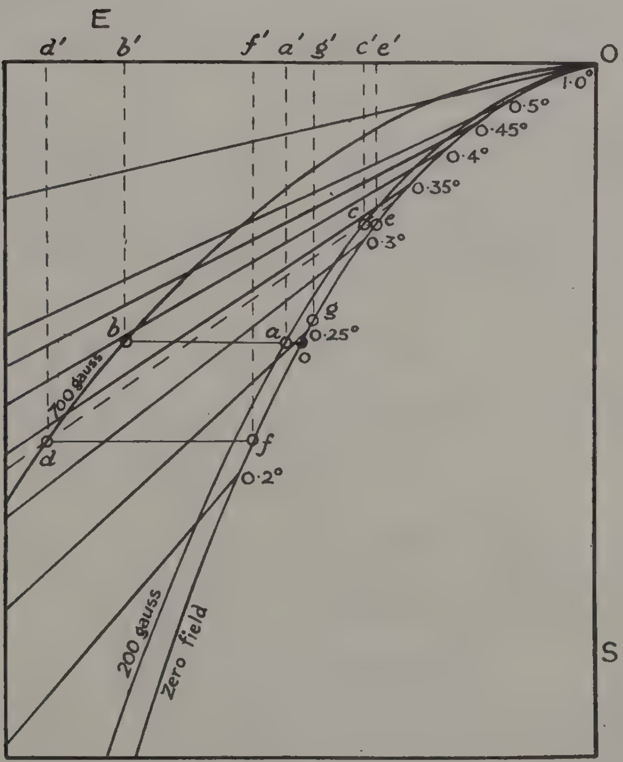
* In principle it would have been possible to do this during the course of the first heat flow, after applying the non-uniform field: in practice, however, fluctuations in such a field, to which the measuring system is very far from astatic, render accurate measurement impossible at this stage.

§ 3. THERMODYNAMICS OF THE TEMPERATURE GRADIENT INITIATION.

The principle of the device for establishing the temperature gradient may best be understood in terms of the Mollier diagram for the salt (fig. 1), in which the entropy S is plotted against the enthalpy E, defined so that $dE = T\,dS - M\,dH$. For the purpose of constructing such a diagram it may be assumed that the susceptibility follows a Curie law $\chi = \lambda/T$ and that the specific heat is of the form $C = A/T^2$, as is certainly a reasonable approximation throughout the temperature range considered. It may then be seen that "lines of constant field" are parabolæ, each one given by allotting to H a suitable value in the equation

$$S \simeq \frac{E^2}{A}\left(1 - \frac{\lambda H^2}{A}\right). \qquad (1)$$

Fig. 1.



"Lines of constant temperature", on the other hand, are straight, being given by the equation

$$E = 2TS \qquad (2)$$

in which the appropriate values are given to T. These lines must necessarily stop at the parabola corresponding to H=0.

Initially the crystal is assumed to be in a condition represented in fig. 1 by the point o. The initial magnetization in the non-uniform field is carried out isentropically, the two ends of the crystal reaching point such as a and b, in different fields and at different temperatures. In the subsequent heat flow the total enthalpy remains unchanged, so that the ends will reach points c and d on the same isotherm, determined by the condition that $a'c'=b'd'$. (These considerations should strictly be applied to all pairs of corresponding points in the two halves of the crystal.) On removing the non-uniform field isentropically the ends move to points e and f, representing temperatures respectively higher and lower than that of the initial state o. Having achieved such a gradient in the presence of zero magnetic field, the subsequent rearrangement of temperature may be studied experimentally and the diffusivity deduced.

After the second heat flow the final state of the system is represented by the point g, fixed by the condition that $e'g'=f'g'$. It may be verified that the temperature at g is somewhat higher than that at O, corresponding to a second-order increase in the entropy consequent upon the irreversible processes which have occurred.

To calculate the order of magnitude of the effect and the exact distribution of temperature set up, we may proceed analytically, using the arguments outlined qualitatively above. If x represents the axial distance between any point in the crystal and one end, the total length being $2l$, and if $H(x)$ represent the field at that point, then it may be shown that the final temperature T' at that point is given in terms of the initial uniform temperature T by the expression

$$T'=T\left\{1-\frac{\lambda}{2A}\left[\{H(x)\}^2-\frac{1}{2l}\int_0^{2l}\{H(x)\}^2\,dx\right]\right\}. \quad . \quad . \quad . \quad (3)$$

If the field $H(x)$ varies uniformly along the length of the crystal following the equation $H(x)=H_0+hx/l$, it being assumed that $h \ll H_0$, then

$$T' \simeq T[1-\lambda H_0 h(x-l)/lA]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The temperature gradient produced is therefore uniform. For potassium chrome alum $\lambda/A=1{\cdot}36 \,.\, 10^{-6}$ gauss^{-2} (Garrett 1948). If $H_0=300$ gauss and $h=100$ gauss, then the extreme difference of temperature produced at 0·2° should be, according to this formula, 0·008°.

§ 4. Experimental Details.

The specimen was a single crystal of chrome alum, 65 mm. in length and of octagonal cross-section, the distance between opposite faces being about 15 mm. The crystal was cut in such a way that the axis of the sample coincided with one of the cubic axes of the crystal. It was suspended by fine nylon threads from a framework of german silver, which itself fits inside the conventional german silver salt-tube of a magnetic cooling cryostat. The measuring coils were wound on paper formers over the outside of the salt-tube, arranged so as to provide an estimate of the averaged susceptibility over the last 15 mm. at each end of the crystal. Measurements of the two mutual inductances could be made alternately by comparison with a standard inductometer using a 40 cycle A.C. bridge network, detection of the balance point being made with the help of an amplifier and vibration galvanometer. In this way a temperature difference on the Curie scale of 0·01° at 0·2° could be determined with an accuracy of about 5 per cent.

The initial cooling was attained with a water-cooled armoured solenoid, which gives a maximum field of 10 Kg. After demagnetization, the non-uniform magnetic field required for setting up the temperature gradient was produced by (i) a solenoid of internal diameter 120 mm. and field-factor 36 gauss/amp.; and (ii) a set of coils fitting inside the solenoid and around the liquid nitrogen dewar, designed to produce a linearly varying field of 10 gauss/amp./cm. A current of 8 amp. passed through these two coils in series for an appropriate interval of time was found to set up a temperature gradient of a suitable magnitude.

After switching off this non-uniform field, measurements of mutual inductance were made for the two coil systems alternately, the time of each determination being noted. Both sets of measuring coils were calibrated at temperatures in the helium range, so that mutual inductance readings could be used in each case to calculate T^*. Unfortunately the corrections to T^* for the shape of the crystal are unknown, and this uncertainty, although still moderately small at the lowest temperature investigated, will be a serious drawback in any attempt to extend the method to lower temperatures. As the field due to the primary of each of the mutual inductances is effective over roughly a spherical volume 15 mm. in diameter near the end of the crystal, it has been supposed that the measurements will give T^*_{sphere} directly.

It may be shown (see Appendix) that, to a good approximation, the difference in temperature between the ends of the crystal falls off exponentially with time according to the law

$$\Delta T \propto \exp\left(-\kappa\pi^2 t/4l^2\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

in which, as before, $2l$ is the overall length and κ the diffusivity of the crystal. In practice there will be, superimposed on these changes of temperature, a steady upward drift due to " heat leak ", which probably represents a surface recondensation of gaseous helium. Smoothed curves representing the terminal temperatures are drawn; from this graph another curve is plotted, showing the difference in temperature plotted logarithmically against the time. The linearity of this graph checks the validity of the exponential law (equation (5)) and the diffusivity of the crystal at a mean temperature may be estimated from the gradient of the fitted straight line. The relation between T^* and the true temperature T is known from Bleaney's measurements (Cooke 1949), as is also the specific heat at each temperature. The thermal conductivity itself may then be derived from the experimental values of the diffusivity.

§ 5. RESULTS.

Measurements were made during three separate helium runs in October 1948. The results are given in the accompanying table.

October	T*	T	$4\kappa^2/l\pi^2$ seconds	C/R	K erg units	$\frac{10^{-5}\mathrm{K}}{\mathrm{T}^3}$
8	0·163	0·145	574	0·693	1·58	5·2
6	0·167	0·149	478	0·682	1·86	5·6
20	0·182	0·166	372	0·635	2·22	4·9
20	0·232	0·218	112	0·475	5·52	5·3
20	0·247	0·235	74	0·427	7·56	5·8
20	0·252	0·240	74	0·405	7·15	5·2
20	0·260	0·248	67	0·395	7·68	5·1
20	0·276	0·266	42	0·358	11·2	5·9
8	0·285	0·274	46	0·340	9·6	4·7

It will be seen that, as expected, the thermal equilibrium time does in fact depend very sharply on temperature. This, incidentally, explains why it has not so far been possible to extend the measurements to temperatures below 0·14° (where the state of transport of heat is no longer large in comparison with the heat leak) or above 0·30° (where an automatic recording device would have been necessary). The results are plotted against temperature in fig. 2, which shows that the conductivity is proportional to T^3 in this temperature range. It may be seen from the table that the average value of K/T^3 is

$$5{\cdot}3\,.\,10^5 \text{ ergs sec.}^{-1}\text{ cm.}^{-1}\text{ degree}^{-1} \quad . \; . \; . \; . \; . \quad (6)$$

with a mean deviation of $0{\cdot}12\,.\,10^5$. Owing to the possibility of systematic errors the actual uncertainty is rather greater than this.

Fig. 2 also shows the results of measurements made by Bijl (1949) in the liquid helium range, where the conductivity was found to follow a $\mathrm{T}^{2\cdot3}$ law. Bijl found that the absolute magnitude of the conductivity appeared to depend on the rapidity with which the crystal was cooled from liquid nitrogen to helium temperatures. No attempt has been made to alter this rate of cooling during the present measurements, but it seems that our conditions correspond more closely with Bijl's " slow cooling ", and it will be seen from fig. 2 that a comparison of the two sets of measurements tends to confirm this. Kürti, Rollin and Simon (1936) estimated the thermal conductivity at 0·18° to be of the order of 10^{-4} watt units, and this also is in agreement with our results.

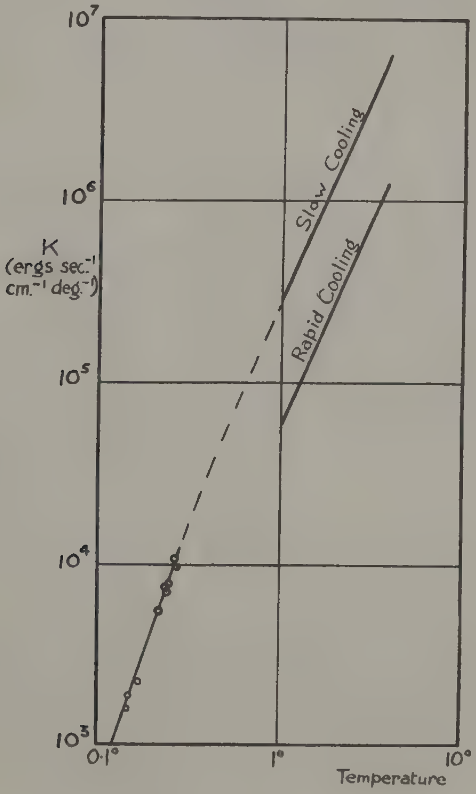
§ 6. DISCUSSION.

In an insulating crystal the principal mechanism of thermal conduction is that associated with Debye waves or phonons. The phonon conductivity is given in order of magnitude by the expression

$$\mathrm{K}_f \sim l_f s \mathrm{C}_f, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (7)$$

C_f being the lattice specific heat per unit volume, s the velocity of sound in the crystal and l_f the phonon mean free path. In this temperature range l_f should be a length specifying the separation between lattice boundaries, and so should be independent of temperature; s is also independent of temperature, and, since the lattice specific heat C_f should be proportional to T^3, it would be expected that the phonon conductivity would also be proportional to T^3.

Fig. 2.



In § 5 it was shown that a T^3 conductivity law is indeed followed for chrome alum at temperatures of the order of 0·2° K., although in the liquid helium range $K \propto T^{2\cdot3}$. Taking $C_f \sim 10^2 T^3$ ergs/° (Casimir 1940) and $s \sim 10^5$ cm./sec. we find that

$$l_f \sim 0{\cdot}05 \text{ cm.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

a length considerably smaller than the macroscopic dimensions of the crystal (6·5 cm. × 1·5 cm.). Probably this is due to a mosaic structure in the crystal, as suggested by de Haas and Biermasz (1938) to account for the "saturation" effect observed by them in experiments on the size dependence of the thermal conductivity of potassium chloride rods. Such a domain structure may also be associated with hysteresis effects in microwave experiments observed by Bleaney and Penrose (1948), and with the observation by Kraus and Nutting (1941) of a shattering of the crystal structure noticed on cooling a very similar material (ammonium chrome alum) through the temperature range 80–100° K.

Akheiser and Pomeranchuk (1944) have made the interesting suggestion that at the lowest temperatures the conductivity of paramagnetic crystals should again become large, due to the alternative mechanism for the transfer of thermal energy provided by the spin system ("exciton conductivity"). This is not in disagreement with the results presented in this paper, for calculation shows that at 0·14° K. the exciton conductivity should be some 5000 times smaller than the phonon conductivity. At somewhat lower temperatures the thermal equilibrium times become exceedingly long, but, if the predictions of Akheiser and Pomeranchuk are indeed correct, the diffusivity at 0·003° K. should be equal to that at 0·14° K., while at still lower temperatures it ought to increase as $1/T^2$. In view of the technical interest of such a possible increase in conductivity it is obviously of considerable importance that experiments be extended to this lower range of temperature.

From the information now available, it is possible to arrive at a tentative conclusion concerning the lowest temperature to which it is possible in a reasonable time to cool a sample of material by thermal contact with a cooled mass of chrome alum. It may be seen from equations (5) and (7) that the thermal equilibrium time within a crystal is proportional to $T^{-3}l^2l_f^{-1}$, where T is the temperature, l the macroscopic size of the sample and l_f the phonon mean path. In an actual experiment (see table of results) an equilibrium time of 500 seconds was found when $l \sim 65$ mm. and $T=0\cdot15°$, l_f being subsequently deduced as 0·5 mm. From these figures it is possible to calculate what would be the equilibrium time at that temperature under the conditions of an actual contact cooling experiment; and hence to deduce how low the temperature could be before the equilibrium time would become intolerably long. In the experiments of Mendoza (1948) in which powdered salt was compressed around a copper insert under considerable pressure, an acetone-based lacquer being used as bonding agent, it was possible to reduce the distance between any grain of salt and the nearest copper surface to about 2 mm.; but on the other hand the phonon mean free path is probably given now by the size of the crystal grains, a distance very much smaller than the phonon mean free path of 0·5 mm. found for a large crystal. As a basis for comparison put $l=1$ mm. and $l_f=5\cdot10^{-3}$ mm. Using the above analysis it is found that at 0·05° K. the equilibrium time for such a sample should be of the order of 1000 seconds, which is about the longest time tolerable in a contact-cooling process. It may therefore be concluded that a temperature of 0·05° K. probably represents the limit to which it would be practicable to cool a sample by direct thermal contact with a cooled salt. In this calculation no account has been taken of any surface thermal resistance between the salt and the metal. This assumption is based on the conclusion of Mendoza (1948) that below 0·2° K. the surface resistance is not the controlling factor.

§ 7. ACKNOWLEDGMENTS.

I should like to thank Professor J. F. Allen and Dr. J. Ashmead for their help during the course of the work described in this paper. Thanks are also due to Mr. F. E. Smith of the University Chemical Laboratory, Cambridge, who supplied the crystal, and to the Department of Scientific and Industrial Research for a maintenance grant.

APPENDIX.

The exact solution of the thermal conduction problem, taking into account a diffusivity varying as T^5, would involve elaborate numerical computation. It is sufficient however to suppose the diffusivity constant at some mean value, and the problem is then soluble by standard methods (Carslaw and Jaeger 1947). Since radiation from the sides is of course negligible, the crystal may be treated as an infinite slab between two parallel planes, there being no heat flow across either boundary. If the initial temperature distribution be $f(x)$, then the temperature at some later time t at any point x is given by

$$T=\frac{1}{2l}\int_0^{2l} f(x')\,dx'+\frac{1}{l}\sum_{n=1}^{\infty} e^{-\frac{\kappa n^2\pi^2 t}{4l^2}} \cos\frac{n\pi x}{2l}\int_0^{2l} f(x')\cos\frac{n\pi x'}{2l}\,dx', \quad . \quad (9)$$

where κ represents the diffusivity of the material. Suppose a linear initial temperature distribution of the form $f(x)=T_0+s(x-l)$. Substituting in equation (9), all terms under the summation sign for which n is even vanish, and there remains the expression

$$\begin{aligned} T=T_0-\frac{8sl}{\pi^2}\cos\left(\frac{\pi x}{2l}\right)\exp\left(-\frac{\kappa\pi^2 t}{4l^2}\right) \\ -\frac{8sl}{9\pi^2}\cos\left(\frac{3\pi x}{2l}\right)\exp\left(-\frac{9\kappa\pi^2 t}{4l^2}\right) \\ -\ldots\ldots \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10) \end{aligned}$$

It is possible, however, to neglect all except the first of the exponential terms, provided that we are not interested in the distribution at times very shortly after the initiation of the heat flow. Since the coefficient of t in the exponent in equation (10) is the same at all points along the length of the crystal, it is immaterial whether temperature measurements are made over an infinitesimal region at the very end of the crystal, or whether instead an average is observed of the temperature near the end of the crystal. In either case

$$\Delta T \sim \exp(-\kappa\pi^2 t/4l^2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

so that a graph of log T against t should be a straight line, from the gradient of which κ may be found.

REFERENCES.

AKHIESER, A., and POMERANCHUK, I., 1944, *Journal of Physics*, **8,** 216.
BLEANEY, B., and PENROSE, R. P., 1948, *Proc. Phys. Soc.*, **60,** 395.
BIJL, D., 1949, *Physica*, **14,** 684.
CARSLAW and JAEGER, 1947, *The Conduction of Heat in Solids*, p. 85.
CASIMIR, H. B. G., 1940, *Magnetism and Very Low Temperatures*, p. 87.
COOKE, A. H., 1949, *Proc. Phys. Soc.*, **62,** 269.
DE HAAS, W. J., and BIERMASZ, TH., 1938, *Physica*, **5,** 47, 320, 619.
GARRETT, C. G. B., 1948, *Cérémonies Langevin–Perrin, Paris* (in the press).
KRAUS, D. L., and NUTTING, G. C., 1941, *J. Chem. Soc.*, **9,** 133.
KURTI, N., ROLLIN, B.-V., and SIMON, F., 1936, *Physica*, **3,** 266.
MENDOZA, E. B., 1948, *Cérémonies Langevin–Perrin, Paris* (in the press).

ERRATA.

In the author's paper on *Quantal Aspects of Scientific Information*, which appeared in the March issue (p. 294), the relation between physical scales and proper scales is unfortunately obscured by an error in the example of §4(c).

The last two sentences on p. 294 should read : " We can now give a probability p of just $\frac{1}{2}$ to a statement in the form : x occupies $\sqrt{i}$ equal intervals. The figure of $\frac{1}{2}$ implies . . . . that its terminal is as likely to fall within the $\sqrt{i}$th equal as to fall outside it."

As pointed out in §4(c), the number of *conceptually*-separate occupance-relations, or metron-content, (§4(b)) is not the same thing as the number of *physically-separate* occupance-relations between a magnitude and the scale on which it is *observed* (§4(a)). Logically, i metrons enable us to add i elements to a formal representation of a measurement. The number of occupance-relations on the scale of *physical magnitude* graduated as in §4(a) may be quite different. In the example given, (§4(c)) $i=x^2/(\Delta x^2)$. But the physical scale-unit of x as defined in §4(a) is Δx, and if the probable error is constant, only $x/\Delta x$ equal intervals can be observationally separated on the scale of x, though i conceptually significant intervals can be defined. The probability is 1/2, not that the ith *conceptual* interval be occupied, as stated, but that the $\sqrt{i}$th *physical* scale-unit be occupied. The number of occupance-relations on the physical scale is, infact, of greater practical importance, though it lacks of course some of the invariant features of metron-content.

D. M. MACKAY.

LVII. *The Crystallography and Thermodynamics of Order–Disorder Transitions in Certain Mixed Ketones.*

By Vera Daniel,
British Electrical Research Association*.

[Received March 20, 1950.]

[Plates XX. & XXI.]

Summary.

Previous work by Oldham and Ubbelohde (1939, 1940) on ketones of equal chainlengths and different positions of the ketogroup has shown that some of these compounds have practically identical crystal structures and physical properties; these will be called equivalent ketones. In mixtures of equivalent ketones all constituent ketones have identical carbon chains, and each ketone contains one ketogroup. Thus, mixtures of equivalent ketones differ from each other only in so far as the configuration of the ketogroups affects the bonding of their structure. In the solid state, the system formed by mixtures of different proportions of equivalent ketones is analogous to a crystalline compound, the bonding of which may be altered at will, within certain limits.

In the present paper the thermodynamic stability of different configurations of ketogroups in mixtures of equal amounts of equivalent ketones is treated theoretically. It is assumed the ketogroups contribute to the bonding only by virtue of their dipole moment.

Different mixtures of six equivalent ketones of chainlength eighteen were examined experimentally. They always form one of two structures, one orthorhombic (or monoclinic) and the other hexagonal, the hexagonal one being relatively disordered, as far as the packing of the chains is concerned. Two kinds of order–disorder transitions occur, one of which corresponds to a segregation of the individual ketones into separate phases, while the other one is analogous to the well known transitions in paraffins. The conditions of stability of the different configurations of ketogroups agree with the theory. In particular, the binding energy of a plane array of ketogroups can be derived from transition temperatures, and agrees with the electrostatic interaction energy derived for this array by Fröhlich (1946).

Introduction.

The relationship between the physical properties of substances and their chemical and crystallographic structure is difficult to investigate because their structure cannot be modified at will. A physical property such as, say, dielectric constant or thermal conductivity, depends on a number of factors. The understanding and control of such a property would be easier if these factors could be altered separately and in small steps. However, this is usually not possible because a change from one chemical compound to another generally alters a number of structural features simultaneously, and in rather large steps.

---

* Communicated by Sir Lawrence Bragg.

Systems which permit the alteration of structural features in small steps can arise in two ways. Firstly, there exist series of compounds such that structural features within the series vary in a regular way and in more or less small stages. Such are the series of homologues in organic chemistry, for instance the paraffins, or the groups of analogous inorganic compounds such as the alkali halides. Here the modification of the structure is effected by chemical means and may be restricted to a more or less distinct structural feature, such as the length of a paraffin chain. Secondly, systems with variable structure may be produced by mixing two or more chemical compounds. Here changes can be made indefinitely small, but they are generally not restricted to any one particular structural feature.

In the present investigation both the above methods are applied simultaneously, in that the materials investigated are mixtures of closely related compounds, where both the original compounds and their mixtures have very similar crystal structures. These mixtures form a system which can in some ways be varied in small and quantitatively controlled steps. It may exist in forms of different degrees of order, and the conditions of stability of these forms can be derived thermodynamically, on the basis of certain simple assumptions.

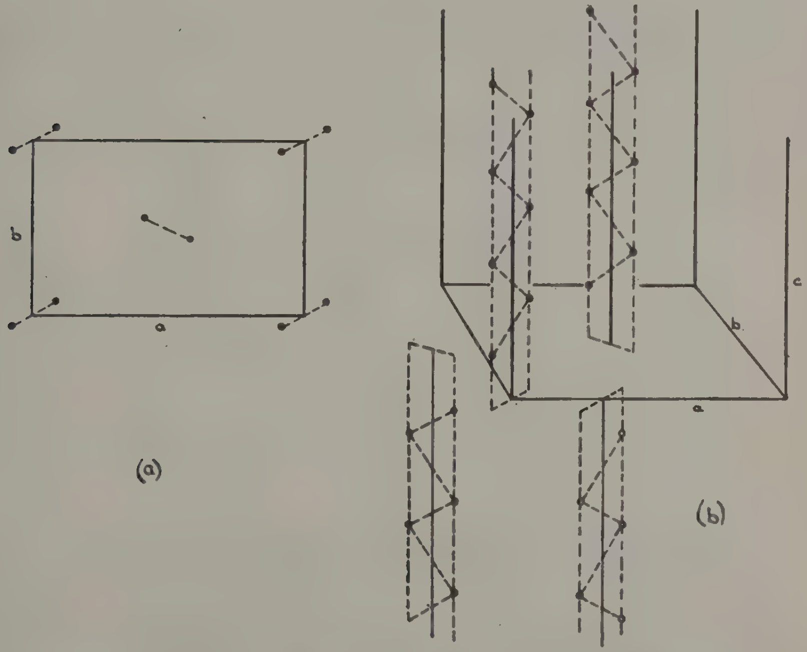
THE STRUCTURE AND PROPERTIES OF LONG-CHAIN KETONES.

The compounds used for the present work are long-chain ketones of the general formula $C_pH_{2p+1}CO\ C_qH_{2q+1}$, with p and q greater than 3. They differ from normal paraffins only in that two hydrogens are replaced by an oxygen. The $C=0$ group has a large dipole moment, and a ketone of this type may be considered as a paraffin to which a dipole has been rigidly attached, approximately at right angles to the chain.

The structure of normal long-chain paraffins has been fairly fully investigated in the past. Müller (1928, 1930) determined the crystal structure of the paraffin $C_{29}H_{60}$ (see fig. 1). This structure is typical for the long-chain paraffins and many of their derivatives. Its main features are that each molecule forms a comparatively rigid flat zig-zag chain and that the chains are parallel to each other in layers. A projection of a layer of chains on the basal plane shows a face-centred arrangement (fig. 1(a)). There exist different modifications of this structure. The chains may be normal or inclined to the basal plane, and the fit of the chain-ends may vary.

Long-chain paraffins show a variety of polymorphic transitions (Piper, Chibnall and Collaborators 1931). Many of these transitions are connected with changes of the angle between the chain axes and the basal plane of the structure, while in other cases the chains are normal to the basal plane throughout the transition. The latter type of transition was investigated by Müller (1932) with the help of X-rays. Müller found that the transition consists in a very anisotropic expansion of the basal plane ab (fig. 1 (*a*)) such that the chains in it approach a hexagonal arrangement. That is the cross section of the chains becomes statistically circular rather than flat. This change of structure is gradual in some paraffins and discontinuous in others. It represents a typical order–disorder transition, as is also confirmed by the calorimetric data of Garner, Bibber and King (1931) and Ubbelohde (1938). Dielectric measurements by Jackson (1935 a, b) Müller (1936, 1939) Pelmore (1939) and Sillars (1939) on paraffins with dipolar groups indicate that in some paraffin lattices the chains have considerable freedom to rotate and/or twist.

Fig. 1.



The structure of the paraffin $C_{29}H_{60}$. (Simplified from Müller 1928.)

As regards ketones as distinct from paraffins, a number of ketones $C_pH_{2p+1}CO\,C_qH_{2q+1}$ can be produced for any one long-chain paraffin of a given chainlength $p+q+1$. Oldham and Ubbelohde (1939) measured the melting points of all ketones of chainlengths seventeen and eighteen, and in addition measured the latent heat of melting and entropy of melting of all ketones of chainlengths seventeen. Table I. gives their results in a slightly simplified form. It shows that for all ketones of the same chainlengths, with *p* and *q* greater than three, all the measured quantities are very similar. The only exception is $C_8H_{17}CO\,C_8H_{17}$ which differs from the other ketones in that its ketogroup is symmetrically placed in the centre of the chain. All the unsymmetrical ketones of equal chainlengths and with ketogroups away from the chain-ends have closely similar latent heats and entropies of melting and melting points.

TABLE I.

Ketones of total chainlength $p+q+1=17$.

p	T_M °C.	ΔH *k* cal./mole	ΔS cal./mole/°
1	47·8	14·1	44·0
2	47·6	13·9	43·2
3	41·6	12·7	40·4
4	41·5	12·7	40·3
5	40·9	12·7	40·5
6	40·6	12·9	40·9
7	41·8	13·1	41·7
8	50·0	13·7	42·3

Ketones of total chainlength $p+q+1=18$

p	T_M °C.
1	51·5
2	51·0
3	45·8
4	45·2
5	44·8
6	45·0
7	44·6
8	45·8

T_M ....melting point,
ΔH ....latent heat of melting,
ΔS ....entropy of melting.

Oldham and Ubbelohde (1939) examined the ketones of chainlengths seventeen by X-rays and microscopically without finding significant differences between them. A similar examination was carried out by the author for ketones of chainlengths eighteen*.

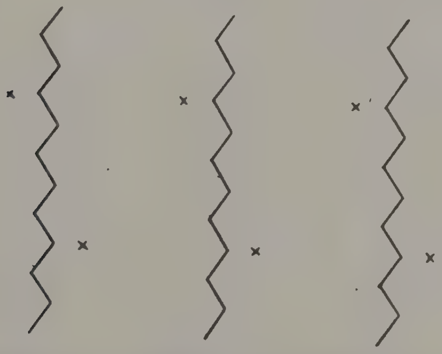
---

* The author is greatly indebted to Dr. Oldham of the Royal Institution for the loan of small samples of five ketones of chainlengths eighteen with $p=3$, 4, 5, 7 and 8. A sixth ketone, with $p=6$, was synthesized according to Dr. Oldham's advice.

X-ray powder photographs were taken of six ketones of that chainlength and $C_6H_{13}COC_{11}H_{23}$ was investigated in somewhat greater detail, using single crystals; but no complete structure determination was made. All the ketones have unit cells of orthorhombic shape, and crystallize with their chain axes normal to the basal plane. (This does not, of course, exclude a monoclinic space-group, and that of $C_6H_{13}COC_{11}H_{23}$ is probably C_{2h}^6 or C_s^4.) No significant difference was found between the unit cell dimensions of the six ketones and inspection of the powder photographs indicates that in all the ketones the chains are arranged very similarly. The X-ray evidence will later be discussed in more detail.

The structure of an unsymmetrical ketone of the kind discussed here has some interesting features which have been considered by Fröhlich (1946)*. As fig. 2 shows, the ketogroups belonging to a layer of chains may be situated at one of two levels. Within either of these two levels

Fig. 2.



Schematic representation of a layer in a crystalline ketone. The positions which may be occupied by oxygen atoms are marked by crosses.

---

* Fröhlich's paper is mainly concerned with a theory of order–disorder transitions in pure ketones. Müller's (1936, 1938) dielectric measurements for the ketones $C_8H_{17}COC_8H_{17}$ and $C_{12}H_{25}COC_{12}H_{25}$ appeared to indicate the inception of such a transition near the melting point: they show the dielectric constant to change gradually, over a temperature range, from the value for the solid to that for the liquid. This gives the impression that a transition would take place at a temperature above the melting point if the material did not melt before this temperature is reached. However, dielectric measurements on very carefully purified samples of $C_8H_{17}COC_8H_{17}$, $C_6H_{13}COC_{11}H_{23}$ and $C_5H_{11}COC_{11}H_{23}$ indicate that in pure solid ketones the dielectric constant is almost independant of temperature, and that there is no gradual transition from the value for the solid to that for the liquid. The change from solid to liquid in a pure ketone is discontinuous at the melting point, and the gradualness of the transition found by Müller appears to be due to small traces of impurities. Thus there is no evidence that pure ketones show order–disorder transitions. The dielectric measurements mentioned will be described elsewhere.

the ketogroups form a planar arrangement, such that the dipoles represented by the ketogroup are roughly parallel to each other. Fröhlich shows that this arrangement is very favourable from the point of view of electrostatic interaction. If a level is fully populated, so that each chain contributes a dipole to it, the electrostatic interaction of the dipoles contributes quite a large term to the binding energy of the structure.

The arrangement of the paraffin chains in the structure leaves it undecided whether ketogroups shall be concentrated in "dipolar planes" or distributed at random between the two possible levels. However, the two possibilities imply different energy conditions. The random arrangement has a higher entropy. On the other hand, the electrostatic interaction is stronger if the ketogroups are concentrated at one level, because the electrostatic interaction energy between dipoles varies inversely with the third power of their distance. If the dipoles are distributed between two levels the numerical value of the total dipolar interaction energy falls to about a quarter of that for a fully populated dipolar plane. If the dipoles are concentrated in dipolar planes the structure will have a lower entropy and a lower energy than if they are distributed at random between two levels, for the dipolar interaction reduces the total energy H of the structure by restricting some of the degrees of freedom. Thus the free energy $G=H-TS$ of the form with dipolar planes has a lower H and a lower S than the random form. Whether or not it has a lower free energy depends on the numerical value of the energy of the dipolar plane. This energy is, as it were, balanced against the entropy term. If it is sufficiently large, the form with dipolar planes will be stable. The thermodynamic conditions of stability in question will be discussed in more detail later.

It is a matter for experiment to decide which of the possible forms of an unsymmetrical ketone is in fact formed. From the point of view of X-ray crystallography the two forms, with and without dipolar planes, differ very little from each other. A preliminary investigation of single crystals of the ketone $C_6H_{13}COC_{11}H_{23}$ showed that it would need a detailed determination of the crystal structure to decide which of the two possible forms is present. However, the two arrangements differ radically in their ability to form solid solutions with other, similar, ketones, as will be seen from the following.

Mixtures of Equivalent Ketones.

As Oldham's and Ubbelohde's work (1939) shows, unsymmetrical ketones of equal chainlengths but different position of the ketogroup have practically identical crystal structure and thermal properties unless their ketogroups are near the chain-ends. These ketones, of the general formula $C_pH_{2p+1}COC_qH_{2q+1}$ with $p\geqslant 3q\geqslant 3$ and $p\neq q$, will be called equivalent ketones. Because of the similarity of the "equivalent" ketones one would expect them to form mixed crystals which were very nearly ideal solid solutions. However, this cannot be the case if the ketones form dipolar planes with a high energy of electrostatic interaction, because in a solid solution these planes must be broken up. Consequently, if the energy of a dipolar plane is large, two equivalent ketones may be mutually immiscible, in spite of their similar structure. On the other hand, if no dipolar planes are formed in the pure equivalent ketones, they should form solid solutions with each other.

Oldham and Ubbelohde (1940) in a study of structures with flaws, measured the melting points of a number of binary mixtures of equivalent ketones. Their results show that in thermodynamic equilibrium these ketones do not mix in the solid to any appreciable extent. This indicates that the pure equivalent ketones crystallize with their ketogroups segregated in dipolar planes.

The two-phase character of binary mixtures of equivalent ketones shows that the energy of dipolar planes must be rather large for the following reason. In pure ketones there exist two levels among which the ketogroups might be distributed at random, while the configuration of the paraffin chains is retained. (See fig. 2.) In a mixture of two equivalent ketones with different positions of the ketogroup there are four such levels, and the formation of dipolar planes means a greater loss of randomness than in a pure ketone. This means that in a mixture of two equivalent ketones the energy of the dipolar planes has to outweigh a larger entropy term if it is to enforce the segregation of dipoles in dipolar planes. The fact that the segregation does occur in binary mixtures implies that the energy of the dipolar planes is sufficient even for that. However, what will happen on mixing three, four, or more equivalent ketones? In these cases there will be still more levels available among which the ketogroups can be distributed at random, and there must be a number of ketones for which the energy of the dipolar plane will be insufficient to enforce the segregation of the dipoles and a solid solution will be formed.

THE THERMODYNAMICS OF MIXTURES OF EQUIVALENT KETONES.

The qualitative reasoning in the preceding paragraph may be made quantitative, subject to some simple assumptions. It may however be useful first to restate the problem in somewhat different terms.

There exist many substances which crystallize at low temperatures in an ordered structure which changes to a disordered structure as the temperature rises. In all these cases an energy term—making for order—militates against an entropy term which makes for disorder. The entropy term contains the temperature as a factor while the energy term does not and so the entropy term must outweigh the energy term at some sufficiently high temperature.

The case of two substances which in the solid state tend to segregate is analogous to an order–disorder transition. Here again an energy term makes for segregation while an entropy term makes for solid solution. Two such substances are said to form a solid solubility gap which closes at some temperature, so that above that temperature a solid solution is always formed.

Mixtures of equivalent ketones have the characteristics both of a single substance undergoing an order–disorder transition and of a number of substances forming a multiple solid-solubility gap. The fact that we have a number of equivalent substances makes it possible to increase the entropy term—which makes for disorder and solid solution—in discrete steps. The disordered form may be made stable by increasing the number of ketones instead of the temperature.

In the following we consider mixtures of equal amounts of n equivalent ketones and deduce their conditions of thermodynamic equilibrium. To this end we make the following assumptions :

1. The ketones considered have equal total energies and entropies for equal temperatures and volumes when in equilibrium.

2. Liquid mixtures of the ketones are ideal solutions.

The first assumption depends for its justification on the equality of the latent heats and entropies of melting of ketones of chainlength seventeen (Oldham and Ubbelohde 1939) and the similarity of the structure of the ketones. The second assumption implies that in liquid ketones the ketogroups are distributed at random.

For solid mixtures of n ketones, Gibbs' phase rule implies that the mixture can consist of not more than n phases in equilibrium (except at a so-called $n-2$ fold point). With regard to the nature of the possible phases we make the assumption :

3. Any ketone in the solid mixture may crystallize either with or without dipolar planes. That is, the ketogroups may either all be concentrated in one level or else distributed at random between all the $2n$ possible levels. This assumption implies that the solid mixture may be either n-phase or single-phase, since if the form with dipolar planes is stable for one ketone, it must be stable for all the ketones. This, because equal amounts of the n ketones are present and thus the specific volume is equal for all of them and because they are equivalent as defined in the first assumption. If the form with dipolar planes is stable the mixture will be n-phase, each phase consisting predominantly of one ketone. If the form without dipolar planes is stable the ketones will form a single-phase solid solution. We shall make one more simplifying assumption with regard to these two possibilities :

4. If the mixture is single phase, it is a completely random solid solution. If it is n-phase, the mass of each ketone is concentrated in its " own " phase, the remainder of its mass which is distributed among the other $n-1$ phases being negligible.

In principle, conditions are likely to exist in which each ketone contains a fair quantity of the other ketones in solid solution, while still forming a separate phase. However, such conditions are considered to be restricted to narrow ranges of temperature so that they may be neglected. This is equivalent to assuming the transition from the form with dipolar planes to the form without dipolar planes to be a transition of the first order.

In a mixture of equal amounts of n equivalent ketones all the ketones behave alike and so it is sufficient to write down the thermodynamic functions for one of them.

For a pure ketone we put the Gibbs free energy per mole in the solid state

$$G_0=H_0-TS_0 \quad . \; . \; . \; . \; . \; . \; . \; . \quad (1)$$

and in the liquid state

$$(G_0)=(H_0)-T(S_0) \quad . \; . \; . \; . \; . \; . \; . \; . \quad (2)$$

the notation used is that of Oldham and Ubbelohde (1940). G is defined for constant pressure so that H contains a volume term.

As the liquid mixture of n equivalent ketones is assumed to be an ideal solution, the partial total energy of each equivalent ketone is the same as for a pure liquid ketone, but the partial entropy is increased by a term equal to $R \log n$ because the volume available for each ketone is increased n fold (see for instance Fowler and Guggenheim, *Statistical Thermodynamics*, p. 161). Thus for the liquid mixture of n ketones

$$(G)=(H_0)-T((S_0)+R\log n). \quad . \; . \; . \; . \; . \; . \; . \quad (3)$$

The solid mixture may be either n-phase or single phase. If it is n-phase the assumptions made imply that all except a negligible part of each ketone is in the same state as in a pure ketone.

Thus
$$G_1=H_0-TS_0 \, . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (4)$$

If the solid mixture is single-phase, the partial entropy of each ketone will be increased by a mixture term, which is again $R \log n$ if the ketones form ideal mixed crystals (see Fowler and Guggenheim, *Statistical Thermodynamics*, p. 185). In addition there will be an entropy increase of $R \log 2$, due to the fact that in a pure unsymmetrical ketone all dipoles are on one level, while in the mixture the dipoles of any one unsymmetrical ketone are on two levels. Apart from this configurational increase of entropy $R \log 2n$ there may be present a further increase of entropy because the removal of the dipolar planes is likely to make the whole structure looser. This dynamic part of the entropy is difficult to estimate. We shall treat it as an unkown constant s, independent of n. This is most likely an over-simplification, as the dynamic part of the excess entropy probably depends on the number of levels on which dipoles may occur, and thus on n, but it seems reasonable to neglect this dependence, in first approximation. The total energy H of the mixture will be increased by the energy of the dipolar plane h_D, and if there is a further loosening of structure beyond the removal of the dipolar planes, it may be increased by a further term h_x. Both these terms are independent of n, in first approximation. This makes the free energy of the single-phase mixture

$$G_2 = H_0 + h_D + h_x - T(S_0 + R \log 2n + s). \quad . \quad . \quad . \quad . \quad (5)$$

The assumptions implicit in this equation mean that the single phase mixture may have another structure than the pure ketone—hence h_x and s—but that the thermal characteristics of this structure are independent of n.

The equations which have been deduced permit calculation of the melting points of the n-phase and single phase mixtures. For the melting point T_1 of the n-phase mixture $G_1 = (G)$, so that

$$(H_0) - H_0 = T_1((S_0) - S_0) + RT_1 \log n$$

where $(H_0) - H_0 = \Delta H$ is the latent heat of melting of a pure ketone and $(S_0) - S_0 = \Delta S$ its entropy of melting. Using the equation $\Delta H = T_0 \Delta S$ valid for the melting point T_0 of the pure ketone, the melting point of the n-phase mixture is given by

$$\Delta H \left(\frac{1}{T_1} - \frac{1}{T_0} \right) = R \log n, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Equation (6) is simply the Hildebrand equation for the melting points of inmiscible substances, written for the special case of equal amounts of substances of equal ΔH and ΔS.

The melting points of the single phase mixtures follow from $(G) = G_2$ and should obey the equation

$$\Delta H \left(\frac{1}{T_2} - \frac{1}{T_0} \right) - \frac{h_D + h_x}{T_2} = R \log 2 + s. \quad . \quad . \quad . \quad . \quad (7)$$

This implies that T_2, the melting point of the single phase mixtures, should be independent of n.

Comparison of equations (6) and (7) shows that with increasing n the melting point of the n-phase mixtures decreases while that of the single phase mixtures remains constant. This implies that even if the multiphase mixture is thermodynamically stable up to the melting point for low values of n, it must become unstable as compared with the single-phase form if n is made large enough. The condition in which the two forms are in equilibrium at a temperature T can be derived by putting $G_1 = G_2$ which leads to the equation

$$h_D + h_x = T(R \log 2n + s). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The lowest value for n at which the single phase form is stable in the solid is obtained if T is the melting point of the single phase form, that is, if $T = T_2$.

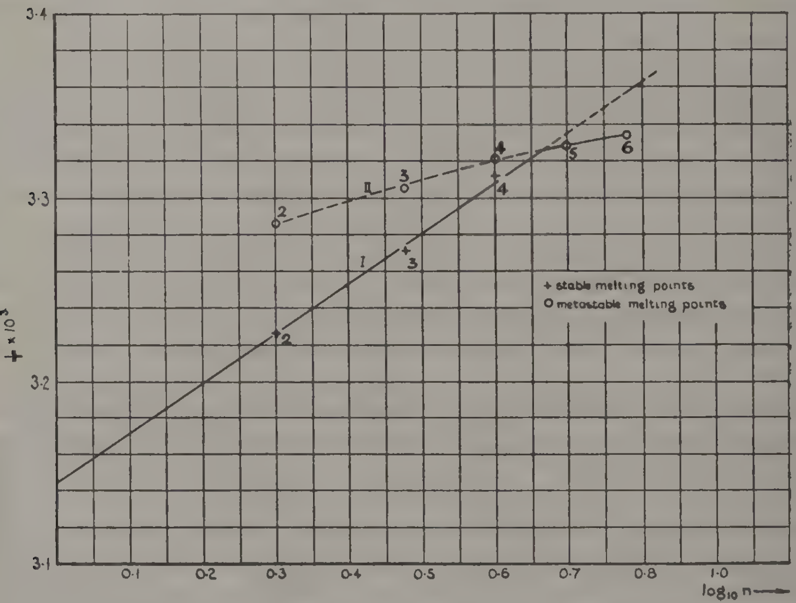
MELTING POINTS OF MIXED KETONES.

Equations (6) to (8) derived in the last paragraph may be compared with experiment by measuring the melting point of mixtures of equivalent ketones. By the courtesy of Dr. Oldham, the author had available small samples of all six equivalent ketones with chainlengths eighteen. Mixtures were prepared of equal amounts of these, and their melting points were measured in the usual way, the mixtures being held in glass capillaries. These measurements led to the observation that mixtures of 2, 3 and 4 ketones can exist in a metastable state. If the mixtures are allowed to solidify slowly, and kept at room temperature for some time, they reach a stable condition to which corresponds a certain melting point. If, however, they are quenched in cold water and then quickly immersed in warm water, they melt at a lower temperature, a metastable melting point. If the liquid mixture is kept for some time at a temperature between the metastable and the stable melting points, it ultimately solidifies again in the stable form. These observations show that, on quenching, the mixtures in question solidify in a mestastable form which changes in the solid state into the stable form if given time. The persistence of the metastable form varies markedly with the number n of ketones in the mixture. Mixtures of four equivalent ketones persist in the metastable state for several days, while in mixtures of two ketones the metastable melting point could only be observed if the quenching in water and ice was followed very quickly by immersion in warm water. No metastable melting points were observed in single ketones.

There exist in general several combinations of equal amounts of n ketones for any number of n. With six equivalent ketones of chainlength eighteen, the number of combinations is, respectively, 15, 20, 15, 6 and 1 for mixtures of 2, 3, 4, 5 and 6 ketones. It would have needed prohibitive amount of material to make all possible mixtures. Three combinations were made of four ketones, and it was found that the values measured for the stable and the metastable melting points did not depend noticeably on the individual ketones making up the four fold mixtures. Variations found were within a range of 1° C. For other mixtures only one or two combinations were used for each value of n.

The results of the measurements are shown in fig. 3. To facilitate comparison with the theoretical equations 6–8, the reciprocals of the melting points are plotted against the logarithm of n. The theory postulates that $1/T$ should vary linearly with $\log n$ if the mixtures are n-phase, while for single-phase mixtures $1/T$ should be independent of n. It may be seen that for $n=1$ to 4 the stable melting points lie on a straight line (I). The metastable melting points depend much less on n and lie on another curve (II), together with the stable melting points of the mixtures of five and six ketones. The experimental results can be interpreted in terms of the theory, if it is assumed that curve I corresponds to n-phase mixtures, while curve II corresponds to single-phase mixtures. This means that up to $n=4{\cdot}6$ where the two curves intersect, the n-phase state is thermodynamically stable while the single phase state is stable for higher values of n (at the melting point). The existence of metastable single-phase mixtures can be easily understood because the formation of a multi-phase mixture from the melt involves extensive sorting out of the mixed ketones. This segregation is a process of diffusion and needs time. Thus, if the mixture is quenched it has to solidify in the random single-phase form. This interpretation explains why mixtures of four ketones persist in the metastable state for several days, while mixtures of two ketones become stable in a few seconds (the temperature being equal in both cases). The segregation of four ketones should need more time than the segregation of two.

Fig. 3.



Melting points of mixtures of n equivalent ketones.

According to equation (6) the slope of curve I should give ΔH, the latent heat of melting of a pure ketone. The value found from the experimental data is $\Delta H = 16{\cdot}6k$ cal/mole. Oldham and Ubbelohde (1940) give the latent heat of melting of one of the equivalent ketones used in the present work, namely $C_5H_{11}COC_{12}H_{25}$, as $\Delta H = 14/6k$ cal/mole. The agreement is satisfactory, as equation (6) embodies some assumptions which are unlikely to be strictly true. The value of $R \log n$ for the entropy of mixing of the liquid mixture of ketones is an upper limit, and this term will be less if assumption 2 is not obeyed and the liquid mixture of ketones is not an ideal solution. If that is the case the value of ΔH deduced from equation (6) is spuriously large. Similarly, it is presumably an exaggeration to assume that the ketones are entirely immiscible in the solid multiphase form. If they are somewhat soluble in each other, equation (6) will again give a spuriously high value of ΔH. As both the factors mentioned tend to make ΔH too large, the agreement found shows that the assumptions made are reasonably correct.

Equation (7) postulates that the melting points of the single-phase mixtures should be independent of n. Experiment shows that these melting points do not depend very strongly on n. However, they are not constant; mixtures of fewer ketones have higher melting points than mixtures of larger numbers of ketones. This is not surprising, for the constancy of h_x, h_D and s in the equations could hardly be expected to be more than a crude first approximation. Some dipolar interaction must occur even in random mixtures and this will be more marked the more densely populated the levels containing dipoles. Dipolar interaction should increase the melting points of single-phase mixtures of few ketones, where the levels are relatively densely populated, in comparison with mixtures of many ketones.

THE CRYSTAL STRUCTURES OF MIXED KETONES.

The evidence of the melting points indicates the existence of two forms of the mixed ketones, a multi-phase form and a single-phase form. It remains now to investigate the structures of these forms, with a view to interpreting the energies and entropies involved in the order–disorder transition.

X-ray powder photographs were taken of a number of mixtures of the six equivalent ketones of chainlengths eighteen, both in the stable and the metastable state. As the most important Debye–Scherrer lines of the ketones occur at Bragg angles of 2 to 20°, an experimental arrangement had to be used which gave sharp lines for these angles. A Unicam single-crystal gonimeter was used, in conjunction with a slit collinator, the slits being about 0·2 mm. wide. The specimens were enclosed in cellophane tubes. For the measurement of spacings flat film was used, the specimen-film distance being 10 cm., and for a survey of the diffraction pattern a cylindrical filmholder of 3 cm. radius.

The results of the X-ray examinations were as follows: In all cases when according to the melting point data an n-phase mixture was expected, its X-ray diffraction pattern was consistent with a multi-phase mixture of the component ketones (see Pl. XX. (b)). In all cases when a single-phase mixture was expected, it had an hexagonal structure, different from that of the pure ketones (see Pl. XX. (a)). The structures of all single-phase mixtures were identical, within the limits of error. In the following these results will be discussed in greater detail.

The six equivalent ketones have closely similar X-ray patterns in that they have identical unit cells and the intensities of most Debye–Scherrer lines are very similar. The reason for this is that all the ketones have identical paraffin chains which are apparently arranged very similarly in the different ketones. As the ketones contain only one oxygen atom for a paraffin chain of eighteen carbon atoms, the intensities of the Debye–Scherrer lines are mostly determined by the paraffin chains. Only if the structure amplitude of the paraffin chain for an X-ray reflection is small does the position of the ketogroup become important for the intensity of that reflection. This happens to be the case for the $00l$ reflections with l up to about 17, and as the $00l$ reflections up to $l=5$ are clearly distinguishable on powder photographs, it is easily possible to identify a ketone from a powder photograph. Otherwise, however, the X-ray diffraction patterns of the ketones are essentially due to their paraffin chains.

The multi-phase mixtures have orthorhombic unit cells, the dimensions of which are closely similar to those of the pure ketones (see Table II. *a*), The signficance of the small discrepancies found is not easy to assess, because, as has been pointed out by McArthur (1944), even with exceedingly pure ketones and paraffins unit cell dimensions are not reproducible within the geometric limits of accuracy. The powder patterns are too complex to be analysed in detail, but as far as visual examination can show, the diffraction patterns of the multi-phase mixtures are very similar to those of the pure ketones with regard to the intensities of the reflections other than the first five $00l$ reflections. This means that the arrangement of the paraffin chains in the multi-phase mixtures is very similar to that in the pure ketones. This agrees with the prediction that these mixtures should be multi-phase mixtures of the pure ketones.

As for the postulate of the theory that the individual ketones in the multi-phase mixtures should form separate aggregates with dipolar planes, the X-ray data are insufficient to give a conclusive answer. The only reflections which are different for the individual ketones, and could be identified on powder photographs, are the first five $00l$ reflections. The intensities of these reflections are only little dependent on whether the individual ketones form separate aggregates. However, even if more data were available, the investigation of this problem would meet the following difficulty.

In general, multi-phase aggregates of crystals may be expected to scatter X-rays incoherently, because the regularity of the repeat of the crystal structure will normally be interrupted at the boundaries between aggregates. However, this need not be the case with a multi-phase mixture of equivalent ketones, where the individual ketones have identically arranged paraffin chains. Aggregates of the different ketones, each containing dipolar planes, may be packed together so as to scatter coherently, as if the whole conglomerate were one single crystal. If this were the case, only a very detailed X-ray investigation could decide whether dipolar planes are formed or not. Such an investigation would be very interesting from the point of view of X-ray optics, but has so far not been undertaken.

TABLE II. *a*.

Cell dimensions of mixtures at room temperature

Mixture	*n*	*a*	*b*	*c*	Remarks
$C_3+C_4+C_5+C_6+C_7+C_8$	6	8·22±0·04	4·74±0·04	25·16±0·15	Stable, hexagonal
$C_3+C_4+C_5+C_6+C_7$	5	8·20±0·04	4·73±0·04	24·76±0·15	,, ,,
$C_4+C_5+C_6+C_7$	4	8·18±0·04	4·72±0·04	24·68±0·15	Metastable, hexagonal
$C_4+C_5+C_6+C_7$	4	7·60±0·04	4·92±0·04	24·94±0·15	Stable, orthorhombic
$C_3+C_5+C_6+C_7$	4	7·60±0·04	5·01±0·04	24·82±0·15	,, ,,
$C_3+C_5+C_6$	3	7·60±0·04	4·89±0·04	24·88±0·15	,, ,,
C_5+C_6	2	7·60±0·04	4·89±0·04	24·92±0·15	,, ,,
Average of six pure ketones	1	7·59±0·02	4·87±0·02	24·78±0·06	,, ,,

TABLE II. *b*.

Mixture	*n*	*a*	*b*	*c*	Remarks	T (°C.)
$C_3+C_4+C_5+C_6+C_7+C_8$	6	7·64±0·04	4·95±0·04	24·68±0·15	Stable, orthorhombic	4
		7·66±0·04	4·97±0·04	24·83±0·15	,, ,,	11
$C_3+C_4+C_5+C_6+C_7$	5	7·56±0·04	4·98±0·04	24·68±0·15	,, ,,	7·4
$C_4+C_5+C_6+C_7$	4	7·56±0·04	4·95±0·04	24·94±0·15	Metastable orthorhombic	7°
		7·54±0·04	5·00±0·04	24·82±0·15	Stable, orthorhombic	6°
$C_3+C_5+C_6+C_7$	4	7·56±0·04	4·80±0·04	25·16±0·15	,, ,,	7°

NOTE: *a*, *b* and *c* in Angstrom units. C_ρ indicates a ketone of formula $C_pH_{2p+1}1\,COC_{17-p}H_{2(17-p)+1}$. Mixtures consist of equal amounts of the *n* ketones.

The above considerations show that the X-ray evidence on the multi-phase mixtures is not quite conclusive. However, the evidence of the unit cell dimensions and the arrangement of the paraffin chains confirms the multi-phase character of these mixtures, and the intensities observed are consistent with it. This appears to establish a strong case for the conclusion drawn from the melting point data, that the stable mixtures of up to four ketones consist of aggregates of the individual ketones, each forming dipolar planes.

As for the single-phase mixtures, X-ray photographs were obtained of stable mixtures of five and six ketones and of metastable mixtures of four ketones ; metastable mixtures of three and two ketones were not photographed, because of their short life. All these mixtures gave X-ray patterns corresponding to a structure with an hexagonal arrangement of chains in the basal plane ab. The structure of the pure ketones is orthorhombic (or alternatively monoclinic with a monoclinic angle of very nearly 90°) and the chains in the basal plane $a\,b$ are arranged in a distorted hexagon (see fig. 1(a)). For the single-phase mixtures the distortion of this hexagon vanishes so that $a=\sqrt{3}(b)$.

The unit cell of the single-phase form may be referred to hexagonal indices, when the reflections 110 and 200 of the orthorhombic notation coalesce into one reflection, with indices 100 in the hexagonal notation. This reflection which is shown on Pl. XX.(a), and another one, corresponding to the coalescing of the orthorhombic 310 and 020 into the hexagonal $2\bar{1}0$ were the only $hk0$ reflections which could be distinguished on the powder photographs. Even the second line, the hexagonal $2\bar{1}0$, was so weak that monochromatic radiation had to be used for its detection. This indicates strong thermal vibrations at right angles to the chains. However, the reflections present establish a hexagonal packing of the chains in which each is surrounded by six nearest neighbours—the closest packing for cylindrical rods.

The $00l$ reflections of the single-phase mixtures correspond to a c axis equal to that of the pure ketones and multi-phase mixtures (see Table II.a). This implies that the paraffin chains are at right angles to the basal plane, as in the pure ketones. The intensities of the $00l$ reflections of any single phase mixture of four ketones are closely similar to those of the multi-phase mixture of the same four ketones. As has been pointed out above, the intensities of these reflections cannot be used to determine whether the ketogroups are distributed at random or arranged in dipolar planes. It may be noted in this context that the high orders of the $00l$ reflections with $l=18$–20 remain visible on the pattern of the single-phase mixtures. These reflections correspond approximately to the so-called backbone spacing of the paraffin chain, 1·27A. The presence of these reflections indicates that thermal vibrations are less violent in the direction of the chains than normal to them.

The hexagonal structure of the single-phase mixtures is more symmetrical than that of the pure ketones and the volume of its unit cell is about

3 per cent larger than for a pure ketone. The hexagonal form thus represents a disordered packing of paraffin chains. It is analogous to the hexagonal high temperature form which occurs in some paraffins (Müller 1932). Müller explained this structure as being due to a twisting and turning of the chains which makes their cross section statistically circular instead of flat. It is interesting that the single-phase form of the mixed ketones is hexagonal over a wide temperature range, while the paraffin of the same chainlength melts without becoming hexagonal. This may be due to the presence of ketogroups at different levels. The dipole moment of these groups implies that they repel or attract each other, according to their configuration, and this is likely to cause some twisting of the chains even in their state of lowest potential energy.

On the whole, the X-ray data of the mixtures of equivalent ketones are in good agreement with the theory. They show that the single-phase form is more symmetrical and thus less ordered than the multiphase form, being hexagonal, while the multiphase form is orthorhombic or monoclinic. The multiphase form has a structure which is consistent with a physical mixture of the pure ketones. However, it has to be borne in mind that the X-ray data could not establish the presence or absence of dipolar planes. It would be desirable to refine the X-ray investigations of these mixtures, using single crystals.

OBSERVATIONS UNDER THE MICROSCOPE.

Although the mixtures of equivalent ketones were not available as single crystals, for the purpose of the X-ray work, they could be made to grow crystallites distinct enough for investigation under the microscope. The technique used was as follows* :—a small quantity of mixture, of the order 5–10 milligrams, is put on a glass slide and covered with a coverslip. The mixture is then molten and the weight of the coverslip makes it spread in a thin film. If the specimen is now cooled slowly, only one or a few crystals form in the direction of the thickness. The crystals tend to be oriented with their *c* axes normal to the glass plates, the more so the slower the cooling. In any specimen it is generally possible to find some crystals filling most of the thickness and with their *c* axes normal to the plates.

Observation in convergent, polarized light, shows that all single phase mixtures are optically uniaxial, the *c* axis being the unique axis. This is in agreement with the hexagonal structure deduced from the X-ray diffractions pattern. The multi-phase mixtures are optically biaxial, with a small axial angle, and no significant difference can be found between these mixtures and the pure ketones. This agrees with their X-ray diffraction pattern.

---

* The author wishes to thank Mr. Scott Harley of the Burmah Oil Company, Ltd., for pointing out this method.

In parallel polarized light metastable mixtures of four ketones may be observed to change into the stable form. The hexagonal form presents a picture in which dark areas—crystals with their unique *c* axes normal to the plate—alternate with coloured crystals of other orientations (see Pl. XXI. (*a*)). When the hexagonal form changes into the orthorhombic form this picture changes. A pattern appears within the dark areas, which consists of bright birefringent needles (see Pl. XXI. (*b*)). These needles seem to represent minute cracks, more or less in crystallographic directions. The contraction manifested by these cracks is in agreement with the fact that the volume of the hexagonal form is 3 per cent larger than that of the monoclinic form.

A LOW TEMPERATURE TRANSITION IN THE SINGLE-PHASE MIXTURES.

When single-phase mixtures of ketones were observed under the microscope, it was found that at about 15° C. they change into a form very similar to the multi-phase form (see Pl. XXI. (*c*)). This change was observed in detail on a microscope stage which could be cooled, using solid CO_2 (details of this microscope stage will be described elsewhere). It was found that on cooling from room temperature birefringent cracks began to appear at about 14·5–15·5° C. and became more intense as the temperature fell, until their intensity reached saturation round 13·5–14° C. These cracks make a pattern very similar to that formed when a single phase mixture of four ketones changes into its multi-phase form. The only difference is that now the cracks are longer and more distinct. On heating, the cracks vanish again, but at a slightly higher temperature; they begin to close at about 14·5–15·5° C. and the last cracks disappear at about 16–17° C. The transformation is rapid, it follows the change of temperature of the microscope stage without observable delay.

The low temperature transition occurs equally in all single phase mixtures of four, five and six ketones, at the same temperatures. The crystal structure of this form was examined by X-rays using a cooling attachment to a "Unicam" single-crystal goniometer. This attachment will be described elsewhere. Pl. XX. (*c*) shows the X-ray powder pattern of the low temperature form. It is indistinguishable from the pattern of the multi-phase form, also with regard to the higher order lines not shown on the Plate. When the multi-phase form is cooled below 14° C. it remains unchanged, both with regard to its X-ray structure and its appearance under the microscope.

As for the physical significance of the low temperature transition, it might be thought at first sight that it corresponds to a segregation of the ketones into the multi-phase form, for equation (8) shows that at sufficiently low temperatures it is always the multi-phase form which is stable. If a mixture of five or six ketones is, in equilibrium, single-phase at its melting point, it must, in equilibrium, break up into the *n*-phase form below a temperature given by equation (8). However, the low temperature transition cannot correspond to this effect, for two reasons. Firstly, it occurs at the same temperature, whether the mixture concerned consists of four, five or six ketones. This contradicts equation (8). Secondly, the transition is very rapid. This is in contradiction with the observation that it takes several days for a mixture of four ketones to change into the stable form, at 18–20° C. At 15° C. the time for the change should be still longer. The rapidity of the low temperature transition implies that it cannot be due to a lengthy process of diffusion.

As the low temperature transition is not due to a segregation of different ketones, it must be due to a contraction of the hexagonal form, the ketogroups remaining distributed at random. The existence of such a transition is not very surprising because it is known that pure paraffins show transitions of this kind from an almost hexagonal high temperature form to a monoclinic or orthorhombic low temperature form (Müller 1932). A mixture of ketones without dipolar planes should not be very different from a paraffin if we assume that the chief contribution of the ketogroup to the binding of the structure is their dipolar interaction. Indeed, the single phase mixtures of five and six equivalent ketones have melting points of 27·5 and 27·0 respectively, as compared with 28° C. for the paraffin $C_{18}H_{38}$. However, the comparison is complicated by the fact that the paraffin $C_{18}H_{38}$ has a structure in which the chains are inclined to the basal plane, while they are normal to it for the mixtures of ketones.

Table II.*b* shows the dimensions of the unit cell for the low temperature form of the single-phase mixtures. These agree with the cell dimensions of the multi-phase form, within the limits of error. As far as the intensities are concerned the powder patterns of the single-phase low temperature form and the multi-phase form are indistinguishable. This implies that the packing of the chains is very similar.

The information provided by the X-ray data makes it possible to derive an expression for the free energy of the low temperature single-phase form. The free energy of the hexagonal form (equation (5)) contains an entropy of mixing $\mathrm{R}\log 2n$. This should be the same for the low temperature form, if the ketogroups remain distributed at random between the possible levels. However, there should be no further difference between the entropy of the low temperature form and that of the pure ketone, because the volume and arrangement of the paraffin chains is the same for both. This should imply an equal mobility of the paraffin chains in the first approximation. Thus the entropy term for the low temperature form is $\mathrm{S}_0+\mathrm{R}\log 2n$. As for the total energy, it should differ from that of a pure ketone only in that dipolar planes are not present. Thus the total energy of the low temperature form should be $\mathrm{H}_0+h_\mathrm{D}$, and the free energy

$$\mathrm{G}_3=\mathrm{H}_0+h_\mathrm{D}-\mathrm{T}(\mathrm{S}_0+\mathrm{R}\log 2n). \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The transition from the low temperature form to the hexagonal form may in principle be a transition of the first or second order. The experimental evidence indicates that it occurs over a temperature interval of 1–2° C. and is thus rather sharp. (The difference between the temperature of the transition or heating and cooling appears to be a hysteresis effect). Thus it seems justifiable to treat the transition as one of the first order. If this is done the transition temperature, T_3, can be derived by equating the free energies of the low temperature and hexagonal forms, which gives

$$h_x = T_3 s. \qquad (10)$$

THE ENERGY OF A DIPOLAR PLANE.

Comparison of equation (10) with with equation (8) shows that the two equations contain three unknown variables, namely h_x, h_D and s. Attempts have been made to find a third equation containing these variables, so that they could be determined, but so far without success. However, equations (8) and (10) allow and approximate calculation of h_D, provided some rough estimate may be found for s, because

$$h_D = RT_1 \log 2n + (T_1 - T_3)s. \qquad (11)$$

Here $n=4{\cdot}6$, T_1 is the transition temperature between multi- and single-phase mixtures for $n=4{\cdot}6$, namely 301° K, and T_3 is the transition temperature between hexagonal and low temperature forms, namely 288° K. As T_1-T_3 is small, h_D is little sensitive to the value of s.

It is easy to find an upper limit for s, because it must certainly be less than the entropy of melting for a pure ketone. This has been measured by Oldham and Ubbelohde (1939) for the series of equivalent ketones of chainlength seventeen, and found to be 41·7 cal/mole degree on the average (see Table I.). This value for s gives $h_D < 1310 + 550$ cal/mole. However, the entropy of melting is certain to be higher than the entropy of a transition in the solid, and it is possible to find a closer approximation for s from the data of Ubbelohde (1938) or Garner, van Bibber and King (1931) who measured specific heats and latent heats of transitions for different paraffins. Ubbelohde's data, which have been evaluated by Fröhlich (1949) give a change of entropy, connected with the transition, of about 6 cal/mole degree. Garner's data give an entropy of transition of about 21 cal/mole degree. Assuming a value intermediate between the two alternatives gives

$$h_D = 1500 \text{ cal/mole}$$

which should be a fair first approximation.

The experimental value for h_D may be compared with the energy of electrostatic interaction for a dipolar plane, which has been calculated by Fröhlich (1946) to be

$$\frac{U}{2} = -\frac{1{\cdot}75}{\epsilon_\infty}\frac{\mu^2}{r^3},$$

where μ is the dipole moment of the C=0 group, r the distance between dipoles in the dipolar plane and ϵ_∞ the dielectric constant of the medium between them. The latter is rather an indefinite quantity because the dipoles are not far enough apart to permit the material between them to be considered as a continuum. This means that ϵ_∞ might have any value between unity and the non-polar part of the dielectric constant of a ketone, namely 2—2·5. Evaluation of U/2 with ϵ_∞ indeterminate gives

$$\frac{U}{2}=\frac{2080}{\epsilon_\infty}\text{ cal/mole},$$

and this may be compared with the experimental value $h_D=U/2=1500$ cal/mole. The two quantities become equal if $\epsilon_\infty=1{\cdot}4$, which seems a very plausible value. Thus the experimentally determined energy of the dipolar planes can be accounted for on the basis of dipolar interaction.

CONCLUSIONS.

Ketones of the formula $C_pH_{2p+1}COC_qH_{2q+1}$ with $p\geqslant3$, $q\geqslant3$ and $p\neq q$ were considered as equivalent ketones throughout the present paper. This corresponds to the assumption that all the bonding forces of these molecules are equal except for the presence of a ketogroup at different parts of the paraffin chain. The ketogroup was assumed to influence the bonding only by means of its dipole moment. With these assumptions, the treatment of mixtures of equal amounts of equivalent ketones could be based exclusively on statistical mechanics and classical electrostatics. This model was on the whole surprisingly successful. It was found that, as far as the packing of the paraffin chains is concerned, the mixtures always have one of two structures, of which one is disordered compared with the other one. The conditions of stability of these structures could be derived from the theoretical model, and were found to depend on the configuration of the ketogroups. Two kinds of order–disorder transitions were observed, one corresponding to the closing of a multiple solid-solubility gap and the other analogous to an order–disorder transition in a paraffin.

In considering the results obtained it has to be borne in mind that the crystallographic data available are somewhat crude and need refining by means of single-crystal work. A more detailed crystallographic investigation would presumably reveal minor differences between mixtures, the structures of which were classed as identical in the present paper. Such an investigation might prove interesting from the point of view of X-ray optics. Apart from this, an X-ray study of the transition from a metastable single-phase to a stable multi-phase mixture might provide interesting data on internal diffusion.

The data obtained in the present work on the conditions of the order–disorder transitions permitted the calculation of the electrostatic energy of a dipolar plane, in first approximation, but not the calculation of the other entropies and energies involved in the transitions. It would be interesting to determine those, for instance by the measurement of specific heats. Such measurement would also provide a check on the validity of the assumptions made.

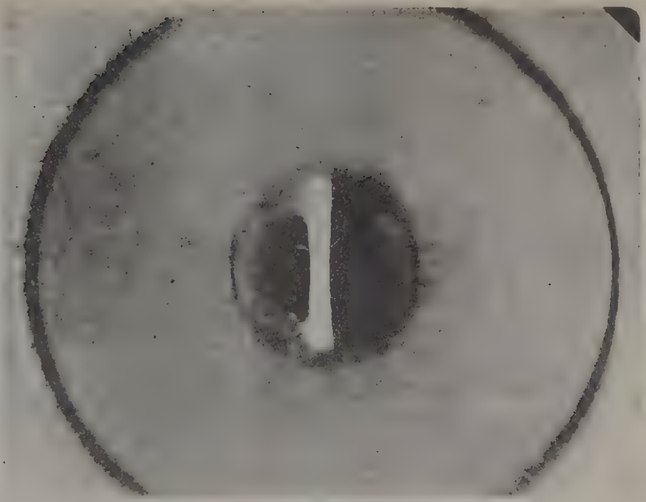
Apart from their intrinsic interest the mixtures of equivalent ketones may provide a tool for studying the dependence of physical properties on structural features in long-chain materials. The different equivalent ketones on the one hand and the various mixtures of ketones on the other hand represent a number of closely related structures, in which structural features vary in different respects. This may prove useful, for instance, for the study of molecular vibrations.

ACKNOWLEDGMENTS.

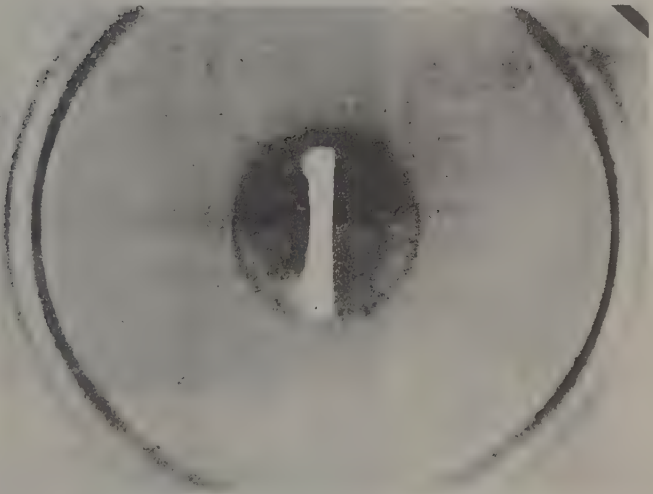
The author wishes to thank Mr. C. G. Garton for valuable suggestions, Dr. J. W. H. Oldham for the loan of samples of ketones and advice on chemical synthesis, and Professor H. Fröhlich and Dr. H. N. V. Temperley for helpful discussions on theoretical aspects of this paper. The author is indebted to the Director of the British Electrical and Allied Industries Research Association for his encouragement and for permission to publish this paper.

REFERENCES.

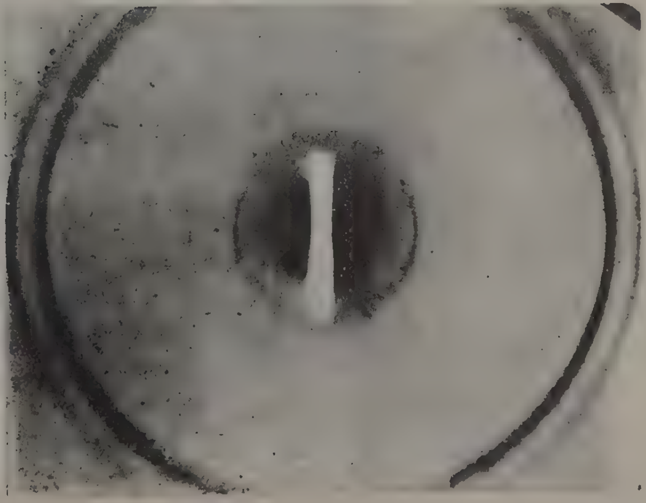
FRÖHLICH, H., 1946, *Proc. Roy. Soc.* A, **185,** 399 ; 1949, *Theory of Dielectrics* (Oxford), p. 148.
FOWLER, R., and GUGGENHEIM, E. A., 1939, *Statistical Thermodynamics* (Cambridge).
GARNER, W. E., VAN BIBBER, K., and KING, A. M., 1931, *J. Chem. Soc.*, 1533.
JACKSON, W., 1935 a, *Proc. Roy. Soc.* A, **150,** 197 ; 1935 b, *Ibid.*, **153,** 158.
MCARTHUR, I., 1944, *Proc. Leeds Phil. Soc.* **4,** 170.
MÜLLER, A., 1928, *Proc. Roy. Soc.* A, **120,** 437 ; 1930, *Ibid*, **127,** 417 ; 1932, *Ibid.*, **138,** 514 ; 1937, *Ibid.*, **158,** 403 ; 1938, *Ibid.*, **166,** 316.
OLDHAM, J. W. H., and UBBELOHDE, A. R., 1939, *Trans. Faraday Soc.*, **35,** 328 ; 1940, *Proc. Roy. Soc.* A, **176,** 50.
PIPER, S. H., CHIBNALL, A. C., HOPKINS, S. J., POLLARD, A., SMITH, J. A. B., and WILLIAMS, E. F., 1931, *Biochem. J.*, **25,** 2087.
PELMORE, D. R., 1939, *Proc. Roy. Soc.* A, **172,** 502.
SILLARS, R. W., 1939, *Proc. Roy. Soc.* A, **169,** 66.
UBBELOHDE, A. R., 1938, *Trans. Faraday Soc.*, **34,** 228.



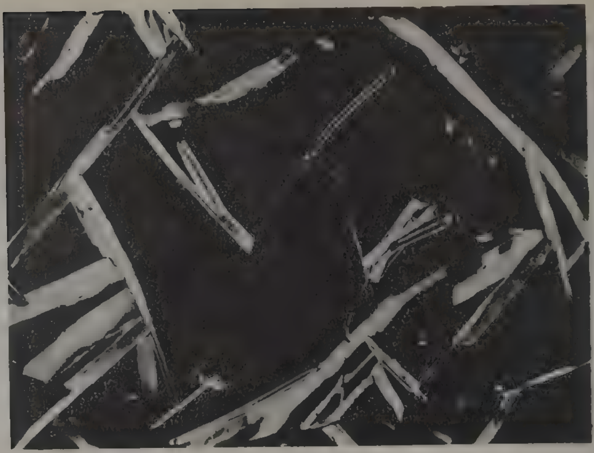
(*a*)
$C_3+C_5+C_6+C_7$
metastable, 18° C.



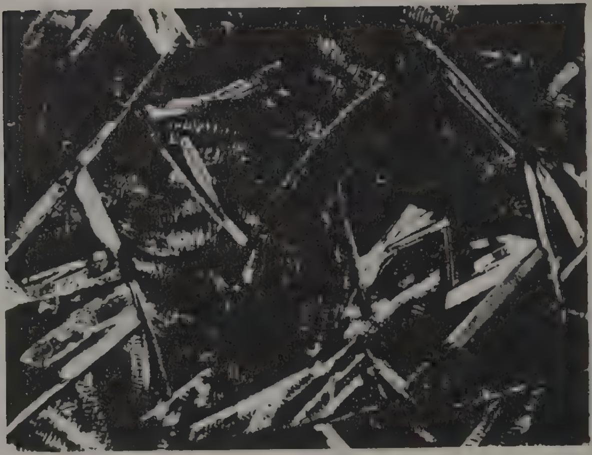
(*b*)
$C_3+C_5+C_6+C_7$
stable, 18° C.



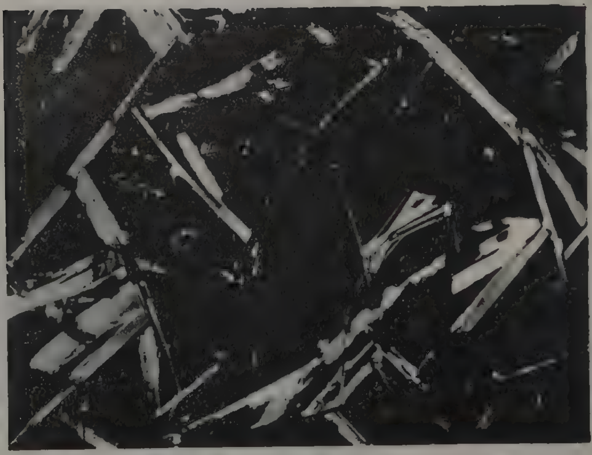
(*c*)
$C_3+C_5+C_6+C_7$
metastable, 11° C.



(*a*)
$C_3+C_4+C_5+C_7$
metastable, r.T.



(*b*)
$C_3+C_4+C_5+C_7$
stable, *r*.T.



(*c*)
$C_3+C_4+C_5+C_7$
metastable, 12° C.

LVIII. *On a Fourth-Order Meson-Equation.*

By W. THIRRING,
Dublin Institute for Advanced Studies *.

[Received January 15, 1950.]

§ 1. INTRODUCTION.

IT was shown recently by the author in a letter to the Editor of the *Physical Review* (Thirring 1950 a), that the quantization of higher order equations leads to results which correspond to the new subtraction method called regularization. But by applying the usual formalism of quantized fields to higher order equations one ought in general to interpret the field as belonging to particles which possess states of several rest masses. One is therefore confronted with the question whether one should accept a realistic or a formalistic interpretation. This question arises in the regularization method as well as in the field theories of Bhabha, Podolsky, etc. It is possible to circumvent this problem by using transcendental operators which possess only one zero-point in momentum space, as was done by Born and Green, but one encounters here the obstacle that a transcendental operator becomes hopelessly difficult, if one tries to introduce generalized derivatives in a space with an electromagnetic field. Another way of avoiding the problem of different rest masses is to take field equations of the form

$$(\Box^2-\mu^2)^n\phi(x)=0.$$

In this paper we shall discuss the next simplest case after the Klein–Gordon equation, namely

$$(\Box^2-\mu^2)(\Box^2-\mu^2)\phi(x)=0.$$

The time independent Green-functuon of this equation is the exponential potential

$$-\frac{\exp(-\mu r)}{2\mu}=\frac{1}{2\mu}\frac{\partial}{\partial\mu}\frac{\exp(-\mu r)}{r}.$$

As this potential has the advantage over the Yukawa potential that it makes no r^{-3} singularity in the tensor force, one is inclined to take this equation for the meson rather than the Klein–Gordon equation. By taking a reasonable meson mass and a simple scalar coupling we can adjust our coupling constant to well known nuclear data, and after having so fixed every constant we may proceed to calculate the self energy of the nucleon due to this meson field. It turns out, that the second order self energy is less than 8 per cent of the rest energy of the nucleus, but that in higher order approximations there are still divergent terms. The reason is that one can make the meson self energy due to virtual nucleon creation finite only by taking another nucleon equation but not by changing the meson equation. The goal of this paper is, of course, not to give a final solution of the divergence problem, but to show that by taking higher order equations some renormalization terms, which used to be infinite, are now finite and have an actual small numerical value. In this paper we use the new formalism of quantized fields, developed by Schwinger and the calculation method of Dyson–Feynman.

---

* Communicated by the Author.

§ 2. FORMALISM.

First we shall define our notation, which is very similar to the one used by Feynman (1949). We choose natural units $\hbar=c=1$ and take real world coordinates with the metric tensor $g_{00}=1$, $g_{11}=g_{22}=g_{33}=-1$. The scalar product of four-vectors we write $a\,b=a_0b_0-\vec{a}\vec{b}$, an arrow indicates a spatial vector. The four-dimensional volume is written without index $dk=dk_0d\vec{k}=dk_0\,dk_1\,dk_2\,dk_3$. By

$$\frac{\partial}{\partial x_\mu} \quad \text{we mean} \quad \frac{\partial}{\partial x_0},\ -\frac{\partial}{\partial x_1},\ -\frac{\partial}{\partial x_2},\ -\frac{\partial}{\partial x_3},$$

the d'Alembertian is

$$\Box^2=-\frac{\partial}{\partial x_\mu}\frac{\partial}{\partial x_\nu}g_{\mu\nu}.$$

The surface-element of a space-like surface is

$$d\sigma_\mu=-dx_1\,dx_2\,dx_3,\ dx_0\,dx_2\,dx_3,\ dx_0\,dx_1\,dx_3,\ dx_0\,dx_1\,dx_2.$$

The nucleons are represented by a quantized Dirac field

$$(i\nabla-m)\psi(x)=0, \qquad \nabla=\gamma_\mu\frac{\partial}{\partial x_\mu}=\beta\frac{\partial}{\partial t}+\beta\,\vec{\alpha}\ \text{grad}.$$

The γ's obey the commutation rules $\gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu=2g_{\mu\nu}$. The scalar product of a four-vector with the matrix vector we write $\mathbf{a}=a_\mu\gamma_\mu$.

The adjoint spinor is given by $\bar{\psi}=\psi^*\gamma_4$, $\bar{\psi}(i\underset{(\leftarrow)}{\nabla}+m)=0$. In momentum space the equations are :

$$\psi=\mathrm{U}_p\exp(-ipx), \quad (\mathbf{p}-m)\mathrm{U}_p=0, \quad \bar{\mathrm{U}}_p(\mathbf{p}+m)=0, \quad p^2=m^2.$$

For the vacuum expectation value we have [Dyson (1949), Feynmen (1949)]

$$\mathrm{P}<\psi(x),\ \bar{\psi}(x')>_0=\tfrac{1}{2}\epsilon(x-x')(i\nabla+m)\Delta_{\mathrm{F}}(x-x')$$

$$=\tfrac{1}{2}\epsilon(x-x')\frac{2i}{(2\pi)^4}\int\frac{\mathbf{p}+m}{p^2-m^2}\exp\{-ip(x-x')\}\,dp$$

$$\epsilon(x)=\frac{x_0}{|x_0|}.$$

P is the operator which arranges factors chronologically. We shall only consider free nucleons with sharp energy-momentum, and get therefore for the one particle expectation value

$$\mathrm{P}<\overline{\psi}(x),\quad \psi(x')>_1=\epsilon(x-x')\overline{\mathrm{U}}_p\mathrm{U}_p\exp\{ip(x-x')\}.$$

Coming now to the meson field we first carry out quantization of the vacuum field and enter then immediately into the interaction representation for the field interacting with the nucleons. From here we can then deduce easily the field equations in the Heisenberg representation. Our vacuum equation is

$$(\Box^2-\mu^2)(\Box^2-\mu^2)\phi_i(x)=0. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The index i indicates the component of the field function in isotopic spin space. Equation (1) is derived from a Lagrangian L

$$\mathrm{L}=-\frac{1}{2}\left(\frac{\partial^2}{\partial x_\varrho^2}\phi_i\frac{\partial^2}{\partial x_\sigma^2}\phi_i-2\mu^2\frac{\partial\phi_i}{\partial x_\varrho}\frac{\partial\phi_i}{\partial x_\varrho}+\mu^4\phi_i\phi_i\right). \qquad . \quad . \quad . \quad . \quad (2)$$

The commutation relations for a field with second order Lagrangian are given by Weiss (1938), Podolsky and Kikuchi (1944), Chang (1946), Chang (1948), De Wet (1948), Le Couteur (1948)

$$\int\left[\frac{\partial\phi_i(x)}{\partial x_0},\quad \frac{\partial\mathrm{L}\{x'\}}{\partial(\partial^2\phi_k(x')/\partial^2x_0')}\right]d\mathrm{V}'=i\delta_{ik}, \qquad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\int\left[\phi_i(x),\quad \frac{\partial\mathrm{L}\{x'\}}{\partial(\partial\phi_k(x')/\partial x_0')}-\frac{\partial}{\partial x_j'}\frac{\partial\mathrm{L}\{x'\}}{\partial(\partial^2\phi_k(x')/\partial x_k'\partial x_j')}\right]d\mathrm{V}'=i\delta_{ik}. \qquad . \quad . \quad . \quad (4)$$

In our case, if we replace the space integration by the integration over an arbitrary space-like surface, we get

$$\int_\sigma\left[\frac{\partial\phi_i(x)}{\partial x_k},\quad \Box'^2\phi_j(x')\right]d\sigma_k=-i\delta_{ij}, \qquad . \quad . \quad . \quad . \quad . \quad (3a)$$

$$\int_\sigma\left[\phi_i(x),\quad 2\mu^2\frac{\partial}{\partial x_k'}\phi_j(x')-\frac{\partial}{\partial x_k'}\Box'^2\phi_j(x')\right]d\sigma_k=-i\delta_{ij}. \qquad . \quad . \quad . \quad (4a)$$

This gives for the commutator function

$$[\phi_i(x),\quad \phi_k(x')]=\delta_{ik}i\mathrm{D}(x-x') \qquad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$\int_\sigma\frac{\partial}{\partial x_\mu}\Box^2\mathrm{D}(x)\,d\sigma_\mu=-1,\qquad \int_\sigma\frac{\partial}{\partial x_\mu}\mathrm{D}(x)\,d\sigma_\mu=0. \qquad . \quad . \quad . \quad . \quad (6)$$

σ is here any space-like surface through the origin. D(x) has to be a solution of (1), therefore the associated function

$$\overline{\mathrm{D}}(x)=\tfrac{1}{2}\epsilon(x)\,\mathrm{D}(x), \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

obeys (1) everywhere but at the origin. To investigate its behaviour there we consider the integral

$$\oint_k dx[\Box^2-\mu^2]^2\,\overline{\mathrm{D}}(x), \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where k is any region round the origin. Applying Gauss' theorem (8) becomes

$$\int_\sigma dS_l \frac{\partial}{\partial x_l}(\Box^2-2\mu^2)\overline{D}(x)+\int_k dx\,\mu^4\overline{D}(x)$$

σ is the surface of k.

If we make the region infinitesimally thin and choose the surface to be space-like we get

$$(8)=\int_\sigma d\sigma_l \frac{\partial}{\partial x_l}(\Box^2-2\mu^2)D(x)=-1.$$

$\overline{D}(x)$ is therefore the invariant Green-function of (1), the same connection with opposite sign holds between the Green-function and the function in the case of the second order field. The invariant Green-function of (1) is

$$\overline{D}(x)=\frac{-1}{(2\pi)^4}\int\frac{\exp(-ikx)}{(k^2-\mu^2)^2}dk=\frac{1}{2\mu}\frac{\partial}{\partial\mu}\overline{D}_\mu(x). \quad . \quad . \quad . \quad (9)$$

$\overline{D}_\mu$ is the Green-function of the Klein–Gordon equation. Though it is evident from the integral-representation one can easily prove that (9) is the right Green-function of (1)

$$(\Box^2-\mu^2)^2\overline{D}(x)=\frac{1}{2\mu}(\Box^2-\mu^2)\frac{\partial}{\partial\mu}(\Box^2-\mu^2)\overline{D}_\mu(x)+(\Box^2-\mu^2)\overline{D}_\mu(x)$$
$$=-\delta(x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

For the commutator function we obtain

$$D(x)=2\epsilon(x)\overline{D}(x)=-\frac{1}{2\mu}\frac{\partial}{\partial\mu}D_\mu(x). \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$D_\mu(x)$ is the D-function for the second order field with rest mass μ and as is necessary, (11) is an odd invariant solution of (1). For the calculation we need Feynman's function which is

$$P<\phi_i(x)\phi_k(x')>_0=\frac{\delta_{ik}}{2}D_F(x)=\frac{\delta_{ik}}{2}\frac{-2i}{(2\pi)^4}\int\frac{\exp(-ikx)}{(k^2-\mu^2)^2}.$$
$$. \quad . \quad . \quad (5a)$$

One notices the similarity between (5*a*) and Feynman's cut-off function.

Passing from the vacuum fields to interacting fields we keep our field equations, but the state vector is now no longer a constant but varies according to the fundamental equation of the interaction representation

$$\frac{\delta\Psi(\sigma)}{\delta\sigma(x)}=-i\mathfrak{H}(x)\Psi(\sigma). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

For the interaction energy $\mathfrak{H}(x)$ we take a simple scalar coupling

$$\mathfrak{H}(x)=-\mu\sqrt{(2)}g_l\phi_l(x)\overline{\psi}(x)\tau_l\psi(x). \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The factor μ is introduced to make the coupling constant g_l dimensionless, τ_l are the isotopic spin operators. The field equations in the Heisenberg representation are easily available. If we indicate the field quantities in the Heisenberg representation for a moment by Clarendon type, the general connection between the derivatives of the field quantitites in the Heisenberg representation and the interaction representation is given by Schwinger (1948)

$$\frac{\partial \mathbf{F}(x)}{\partial x_\nu}=\mathrm{U}^{-1}(\sigma)\frac{\partial \mathrm{F}(x)}{\partial x_\nu}\mathrm{U}(\sigma)+i\mathrm{U}^{-1}(\sigma)\int_\sigma [\mathrm{H}(x), \mathrm{F}(x')]\, d\sigma'_\nu\, \mathrm{U}(\sigma). \tag{14}$$

U is the unitary operator which performs the transformation between the Heisenberg and the interaction representations. Using (6), (13), (14) we obtain

$$\left.\begin{aligned}
&\frac{\partial \boldsymbol{\varphi}_i}{\partial x_\nu}=\mathrm{U}^{-1}\frac{\partial \phi_i}{\partial x_\nu}\mathrm{U}, \qquad \Box^2\boldsymbol{\varphi}_i=\mathrm{U}^{-1}\Box^2\phi_i\mathrm{U},\\
&\frac{\partial}{\partial x_\nu}\Box^2\boldsymbol{\varphi}_i=\mathrm{U}^{-1}\frac{\partial}{\partial x_\nu}\Box^2\phi_l\mathrm{U},\\
&\Box^2\Box^2\boldsymbol{\varphi}_i=\mathrm{U}^{-1}\Box^2\Box^2\phi_i\mathrm{U}+\mathrm{U}^{-1}\mu\sqrt{(2)}g_i\bar{\psi}(x)\tau_i\psi(x)\mathrm{U},
\end{aligned}\right\} \quad . \; . \; . \tag{15}$$

therefore the field equations for the meson field in the Heisenberg representation are :

$$(\Box^2-\mu^2)^2\boldsymbol{\varphi}_i=g_i\mu\sqrt{(2)}\bar{\boldsymbol{\psi}}\tau_i\boldsymbol{\psi}. \quad . \; . \; . \; . \; . \; . \; . \tag{16}$$

To prove the third equation of (15) we need

$$\int_\sigma \Box^2 \mathrm{D}(x)\, d\sigma_\nu=\frac{-1}{2\mu}\frac{\partial}{\partial\mu}\int_\sigma \mu^2 D_\mu(x)\, d\sigma_\nu=0.$$

For the spinor field we get [Schwinger (1948)] if we use

$$\int\{\bar{\psi}_\alpha(x), \quad (\gamma_\mu\psi(x'))_\beta\}\, d\sigma'_\mu=\delta_{\alpha\beta} \quad \{\mathrm{A}, \mathrm{B}\}=\mathrm{AB}+\mathrm{BA},$$

and the identity

$$[\bar{\psi}(x)\psi(x), \psi(x')]=\psi(x)\{\bar{\psi}(x), \psi(x')\}-\{\bar{\psi}(x), \psi(x')\}\psi(x)$$

the following field-equations in the Heisenberg representation

$$(i\nabla-m+g_i\mu\sqrt{(2)}\tau_i\boldsymbol{\varphi}_i)\boldsymbol{\psi}=0. \quad . \; . \; . \; . \; . \; . \; . \tag{17}$$

The field quantities in the Heisenberg representation are easily expressed in terms of the field quantities in the interaction representation [Schwinger (1949 b), Dyson (1949)]. For the interaction energy in the Heisenberg representation we get

$$\mathbf{H}(x_0)=\sum_{n=0}^{\infty}\frac{(-i)^n}{n\,!}\int_{-\infty}^{\infty}dx_1\ldots\int_{-\infty}^{\infty}dx_n\mathrm{P}(\boldsymbol{\mathfrak{H}}(x_0)\boldsymbol{\mathfrak{H}}(x_1)\ldots\boldsymbol{\mathfrak{H}}(x_n)). \tag{18}$$

First we consider the expectation value of the second term of the series (17) for a state where two nucleons and one meson are present, which is nothing but twice* the second order interaction between two nucleons.

---

* Half of it is subtracted if we transform H_0.

We get, if we integrate over the space volume,

$$2\mathrm{E}_{\mathrm{int}}=\int \mathrm{H}_2(x_0)\, d\vec{x}_0$$
$$=-i2\mu g_k g_l \int d\vec{x}_0\, dx_1 \mathrm{P} < \psi(x_0)\tau_l\psi(x_0)\bar{\psi}(x_1)\tau_k\psi(x_1)\phi_l(x_0)\phi_k(x_1) >_{2,\,0}.$$

For simplicity we neglect, in evaluating the expectation value, exchange terms which give rise to mixed densities.

$$\mathrm{E}_{\mathrm{int}}=2\mu^2 g_l^2 \frac{1}{2(2\pi)^4}\,\frac{-1}{2\mu}\,\frac{\partial}{\partial\mu}\int_{-\infty}^{\infty}[\rho_l^1(x_0)\rho_l^2(x_1)+\rho_l^2(x_0)\rho_l^1(x_1)]$$
$$\times\frac{\exp\{-ik(x_0-x_1)\}}{k^2-\mu^2}\, d\vec{x}_0\, dx_1\, dk. \quad . \; . \; . \; . \; . \; . \quad (19)$$

The upper index on the nucleon densities $\rho_l^i(x)=<\bar{\psi}(x)\tau_l\psi(x)>$ indicates to which nucleon the density belongs. To obtain the static energy we assume the densities to be time independent, carry out the time integration and get

$$\mathrm{E}_{\mathrm{int}}=-\mu\frac{g_l^2}{4\pi}\int \rho_l^1(\vec{x}_0)\rho_l^2(\vec{x}_1)\exp\{-\mu(x_0-x_1)\}\, d\vec{x}_0\, d\vec{x}_1. \quad . \; . \quad (20)$$

Or if we assume the densities to be sufficiently concentrated and approximate them by a δ-function

$$\mathrm{E}_{\mathrm{int}}=-\mu\frac{g_l^2}{4\pi}\tau_l^1\tau_l^2\exp\{-\mu(x_1-x_2)\}. \quad . \; . \; . \; . \; . \quad (21)$$

x_1 and x_2 are the positions of the nucleons. The negative sign shows that we get attractive forces. The same result yields a classical meson theory if we start from equations (16) and (13) and replace the nucleon by a point source. We are now able to determine the coupling constant g. We may either choose symmetric theory $g=g_1=g_2=g_3$, $g_4=0$, or neutral theory $g=g_4$, $g_1=g_2=g_3=0$. In both cases $\mu g^2/4\pi$ is the strength of the potential. To fit the proton–proton scattering data for an exponential potential we need a strength of about 100 MeV. if we choose a meson mass of 282 m (Rosenfeld 1948). This gives a coupling constant $g^2/4\pi=0{\cdot}71$. It is clear that the expontial potential wants a stronger coupling than the Yukawa potential with its singularity. Now everything is ready for working out the self energy.

§ 3. CALCULATION OF THE SECOND ORDER SELF ENERGY.

The second order self energy is given by the expectation value of the second term in (17) for a state where one nucleon and no meson is present. The expectation value

$$<\bar{\psi}(x_0)\psi(x_0)\bar{\psi}(x_1)\psi(x_1)\phi(x_0)\phi(x_1)>_{1,\,0}$$

is of course

$$<\bar{\psi}(x_0)\psi(x_0)\bar{\psi}(x_1)\psi(x_1)>_1<\phi(x_0)\phi(x_1)>_0.$$

The one-particle part is Schwinger (1949 a)

$$<\overline{\psi}(x_0)\psi(x_0)\overline{\psi}(x_1)\psi(x_1)>_1=<\overline{\psi}(x_0)\psi(x_1)>_1<\psi(x_0)\overline{\psi}(x_1)>_0$$
$$+<\psi(x_0)\overline{\psi}(x_1)>_1<\overline{\psi}(x_0)\psi(x_1)>_0$$

which makes the self energy

$$\begin{aligned}\mathrm{E}_{\text{self}}&=\int<\mathrm{H}_2(x_0)>_{1,0}\,d\vec{x}_0\\ &=-i2\mu^2 g_l g_k\int d\vec{x}_0\,dx_1<\tau^l_{\alpha\beta}\tau^k_{\gamma\delta}>_1\mathrm{P}<\phi_l(x_0)\phi_k(x_1)>_0\\ &\quad\times[\mathrm{P}<\overline{\psi}_\alpha(x_0)\psi_\delta(x_1)>_1\mathrm{P}<\overline{\psi}_\gamma(x_1)\psi_\beta(x_0)>_0\\ &\quad+\mathrm{P}<\overline{\psi}_\gamma(x_1)\psi_\beta(x_0)>_1\mathrm{P}<\overline{\psi}_\alpha(x_0)\psi_\delta(x_1)>_0].\quad\ldots\quad(22)\end{aligned}$$

Inserting (5*a*) and the expectation values for the spinor field and taking into account that the vacuum expectation value of two spinors gives Kronecker operators in spin and isotopic spin spaces we get

$$\mathrm{E}_{\text{self}}=-i4\mu^2 g_l^2\int d\vec{x}_0\,dx_1\,dp'\,dx\,\epsilon(x_0-x_1)\overline{\mathrm{U}}_p\exp\{ip(x_0-x_1)\}$$
$$\times\epsilon(x_0-x_1)\frac{i}{(2\pi)^4}\frac{\mathbf{p}'+m}{p'^2-m^2}\mathrm{U}_p\frac{-i}{(2\pi)^4}\frac{\exp\{-ik(x_0-x_1)\}}{(k^2-\mu^2)^2}<\tau_l\tau_l>_1.\quad\ldots\quad(23)$$

It is easy to see that the two different terms in the one particle expectation value give rise to the same integral (23) with different sign of k. The sign can be restored by a change of integration variable, we may therefore replace the second term by a factor two in the first one. Further, the square of every isotopic matrix τ is the unit matrix, all kinds of mesons give therefore the same contribution to the self energy, which is equal for neutron and proton. We will now drop the index on the coupling constant and calculate the contribution to the self energy of one kind of meson. The integration over space and nucleon momentum in (23) yields then

$$\mathrm{E}=-i\frac{4\mu^2g^2}{(2\pi)^4}\int\overline{\mathrm{U}}_p\frac{\mathbf{p}-\mathbf{k}+m}{(k^2-\mu^2)^2(k^2-2pk)}\mathrm{U}_p\,dk.\quad\ldots\quad(24)$$

We have integrated over the x_0 unit volume, because we shall renormalize our spinor function to the unit volume. To carry out the operation in spin space which is required in (24) we observe that it is compatible with the Dirac equation to choose the vector $\overline{\mathrm{U}}_\alpha(p)$ in spin space so that $\overline{\mathrm{U}}_\alpha\mathrm{U}_\beta=\mathrm{N}(\mathbf{p}+m)_{\beta\alpha}$. As desired this corresponds to an unpolarized wave since $\overline{\mathrm{U}}\sigma_{\mu\nu}\mathrm{U}=0$. The normalization constant N we may take to normalize either to one mesonic charge per unit volume $\overline{\mathrm{U}}_\alpha\mathrm{U}_\alpha=1$, $\mathrm{N}=1/4m$ or to one electric charge per unit volume $\overline{\mathrm{U}}_\alpha\gamma^0_{\alpha\beta}\mathrm{U}_\beta=1$, $\mathrm{N}=4p_0/1$. For convenience we take the former possibility.

After an easy spur computation the integral (24) is then evaluated by a straightforward calculation using Feynman's method (Feynman 1949) and we obtain

$$\mathrm{E}_{\mathrm{self}}=\frac{-g^2\mu^2}{2(2\pi)^2m^2}\int_0^1\frac{m2x(2-(1-x))}{(1-x)^2+x\mu^2/m^2}dx=-\frac{4g^2}{(4\pi)^2}m\left(\frac{\mu}{m}\right)^2\int_0^1\frac{x^2+x}{(x-x_1)(x-x_2)}dx. \quad . \; . \; . \; (25)$$

The zeros of the denominator in (25) are worked out to be

$$x_{1,\,2}=\exp\,(\pm i\gamma),\quad \tan\gamma=\frac{\surd(2\alpha-\alpha^2)}{1-\alpha},\quad \alpha=\frac{\mu^2}{2m^2}=0\cdot 0118,\quad \gamma=0\cdot 153. \quad . \; . \; . \; (26)$$

Introducing the further abbreviations

$$x_{12}-1=-\alpha\pm i\surd(2\alpha-\alpha^2)=\mathrm{P}\exp\,(\pm i\beta),\quad \mathrm{P}=0\cdot 1535,\quad \beta=1\cdot 647, \quad . \; . \; . \; (27)$$

the elementary integration yields

$$\left.\begin{aligned}\int_0^1\frac{x^2}{(x-x_1)(x-x_2)}dx&=1+\frac{1}{\sin\gamma}\{\sin 2\gamma\,.\,\ln\mathrm{P}+(\beta-\gamma)\cos 2\gamma\}=6\cdot 3\\ \int_0^1\frac{x}{(x-x_1)(x-x_2)}dx&=\frac{1}{\sin\gamma}\{\sin\gamma\ln\mathrm{P}+(\beta-\gamma)\cos\gamma\}=7\cdot 7.\end{aligned}\right\} . \quad (28)$$

Inserting into (25) we get

$$\mathrm{E}=\frac{-g^2}{\pi}\left(\frac{\mu}{m}\right)^2\frac{14\cdot 0}{4\pi}m=-m\frac{g^2}{\pi}0\cdot 025=-m\,0\cdot 076. \quad . \; . \; . \; (29)$$

The negative sign corresponds to the fact that we have attractive forces. For pseudoscalar coupling the self energy is even less, namely

$$\mathrm{E}_{\mathrm{self(ps)}}=-m\frac{f^2}{\pi}0\cdot 00454.$$

The second order self energy gives therefore only a small contribution to the observable mass of the nucleon.

It might be interesting to calculate the classical self energy. If we replace the nucleon density in (16) for the static case

$$(\Delta+\mu^2)^2\phi(x)=\rho(x) \quad . \; . \; . \; . \; . \; . \; . \; . \; (16a)$$

by a point source

$$\rho(x)=g\mu\surd(2)\delta(\vec{x})$$

and use the above mentioned solution of (16*a*) we get for the self energy

$$\mathrm{E}_{\mathrm{self\,(cl)}}=-\frac{1}{2}\int\rho(x)\phi(x)\,d\vec{x}=-\frac{1}{2}\int g\mu\surd(2)\,\delta(\vec{x})\frac{g\mu\surd(2)}{2\mu}\frac{\exp\,(-\mu r)}{4\pi}d\vec{x}$$

$$=\frac{-\mu^2g^2}{(2\pi)^4}\int 2\pi\delta(k_0)\frac{dx\,dk}{(k^2-\mu^2)^2}=\frac{-g^2}{4\pi}m\frac{\mu}{2m}=\frac{-g^2}{4\pi}0\cdot 0768m. \quad . \quad (30)$$

Comparing the integral representation of the classical self energy given in (28) with (24) we see that the self energy of the quantized field in (24) corresponds to a spread out of the nucleon density. This causes the reinforcement of the denominator in (24) which cuts off the contributions of the high frequencies and makes the quantum theoretical self energy a bit smaller than the classical one.

§ 4. HIGHER ORDER APPROXIMATIONS.

In the fourth order contribution to the self energy there are terms which are even smaller than the second order self energy and which correspond to double emission of mesons. The Feynman graphs of these terms would be

Fig. 1.

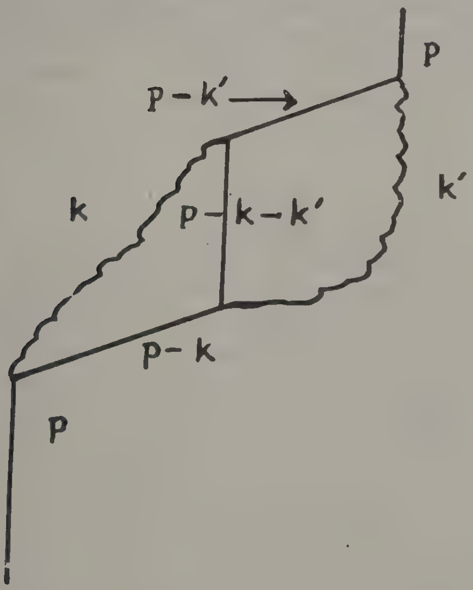
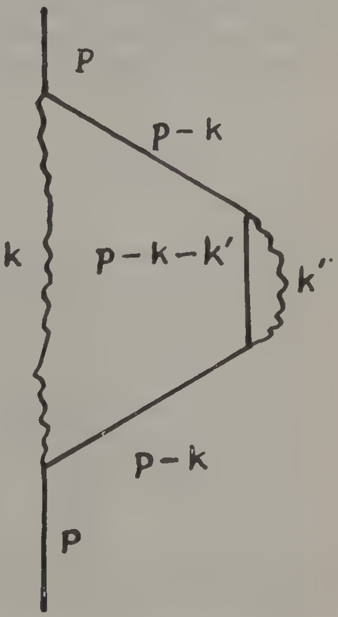


Fig. 2.



but there are still divergent terms arising from the creation of nucleons as in the following graph:

Calculating the contribution to the self energy due to this graph we encounter the following integral :

$$\int \frac{(\mathbf{p}-\mathbf{k}+m)tr(\mathbf{p}'+m)(\mathbf{p}'-\mathbf{k}-m)}{(k^2-\mu^2)^4(p'^2-m^2)((p'-k)^2-m^2)}\,dp'\,dk.$$

Though the denominator is of degree 12 and the numerator is of degree 11, this integral is divergent as it contains the self energy of the meson which has the divergent integral

$$\int \frac{tr(\mathbf{p}'+m)(\mathbf{p}'-\mathbf{k}+m)}{(p'^2-m^2)((p'-k)^2-m^2)}\,dp'.$$

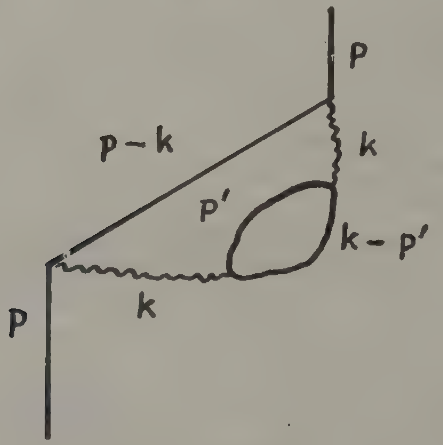
On this integral the form of the meson equation has no influence and it is infinite in our theory as well as in the usual one. The same fact mars the other higher order terms.

In conclusion we may state, that it is possible to make finite terms which are infinite in the usual theory, by taking higher order equations, but it does not yet seem to be possible to select from the manifold of linear equations of arbitrary order just those equations which might be realized in nature.

I should like to take this opportunity of thanking Professor Schrödinger for valuable discussions in connection with this paper.

Note added in proof.—In the meantime the same equation was proposed independently by Bhabha (*Phys. Rev.*, **77,** 665, 1950). For a closer discussion of the difficulties of this theory see Pais and Uhlenbeck (to appear in *Phys. Rev.*, **77,** 1st July, 1950).

Fig. 3.



REFERENCES.

CHANG, T., 1946, *Proc. Camb. Phil. Soc.*, **42,** 133; 1948, *Ibid.*, **44,** 76.
DE WET, J., 1948, *Proc. Camb. Phil. Soc.*, **44,** 546.
DYSON, F., 1949, *Phys. Rev.*, **75,** 486.
FEYNMAN, P., 1949, *Phys. Rev.*, **76,** 769.
LE COUTEUR, K., 1948, *Proc. Camb. Phil. Soc.*, **44,** 63.
PODOLSKY, B., and KIKUCHI, C., 1944, *Phys. Rev.*, **65,** 229.
ROSENFELD, L., 1948, *Nuclear Forces.* North-Holland Publishing Company, (Amsterdam).
SCHWINGER, J., 1948, *Phys. Rev.*, **74,** 1439; 1949 a, *Ibid.*, **75,** 651; 1949 b, *Ibid.*, **76,** 790.
THIRRING, W., 1950 a, *Phys. Rev.*, **77,** 570; 1950 b, *Acta Phys. Austr.* (in the press).
WEISS, P., 1938, *Proc. Roy. Soc.* A, **169,** 102.

LIX. *The Effect of Electron Concentration on the Lattice Spacings in Magnesium Solid Solutions.*

By H. JONES,
Mathematics Department, Imperial College, London *.

[Received April 6, 1950.]

SUMMARY.

The hypothesis that the change in the c/a ratio in magnesium solid solutions is due to the changes in the Fermi energy, arising from the distortion of the Brillouin zone, is developed into a more general and quantitative theory. The case of monovalent solutes is considered, and agreement with experiment is shown to be not unsatisfactory. The experimental results enable deductions to be made regarding the density of electronic states in pure magnesium.

§ 1. INTRODUCTION.

THE changes which occur in the lattice parameters a and c of magnesium solid solutions, as the concentration of the solute is varied, have been measured with great accuracy by Hume-Rothery and Raynor (1940) and Raynor (1942). A simple qualitative theory has already been proposed to account for the variation of the c/a ratio in those solid solutions for which the valency of the solute is greater than that of magnesium (Jones 1949). The experimental results show unambiguously, in all cases, that this c/a ratio is determined, not by the atomic concentration of the solute, but by the valence electron concentration in the alloy.

It is the purpose of this paper to express the theory in a more general and quantitative form than hitherto, and to extend it to include those cases where the valency of the solute is less than the valency of magnesium, *e. g.* to the cases of lithium and silver.

The changes brought about in the overall form of the lattice by the addition of other elements in solid solutions may be regarded as the sum of two effects, (i) a change of atomic volume, and (ii) the development of a shearing strain. The change in the atomic volume must depend upon many factors, *e. g.* the difference between the size of the solute ion and the size of the magnesium ion, and also upon the difference in the potential fields of the solute and solvent ions in which the valence electrons move. The complexity of the causes of the change of volume can clearly be seen in the experimental observations which show no obvious regularities between the dilatation at a given concentration and the solute metal.

---

* Communicated by the Author.

Since the structure of pure magnesium is almost exactly that of hexagonally close packed spheres (the c/a ratio for Mg is 1·6237 and the value for ideal packing 1·6330), and since the metal is comparatively elastically isotropic, there is no reason to believe that differences of ionic size, or of spherically symmetrical potential fields, will play a significant part in producing the shearing strain. It might, therefore, be expected *a priori* that the shearing strain would result mainly from the effect of the Brillouin-zone structure on the valence electrons of the metal. This supposition is confirmed by the fact, already mentioned, that the shearing strain is substantially independent of the particular solute metal and depends, to a large degree, only on the ratio of the number of electrons to atoms in the metal.

§ 2. A CONVENIENT FORMULATION OF THE EXPERIMENTAL RESULTS.

It may be assumed that the relations between the lattice parameters and the concentrations of solute atoms will not be significantly different at the absolute zero of temperature from the observed relations at room temperature. Thus it will be sufficient to express the condition for equilibrium by minimizing the energy rather than the free energy.

The energy per atom W, expressed relative to the energy of the separated atoms, is a function of the atomic volume v, of the shearing strain, and, in the first place, of the solute concentration x. It is convenient to introduce the variable $s=(c/a)^{2/3}$ so that $a=(4v/\sqrt{3})^{1/3}s^{-1/2}$ and $c=(4v/\sqrt{3})^{1/3}s$. The conditions for equilibrium can then be written :

$$\text{(i)}\ \left(\frac{\partial \mathrm{W}(v, s, x)}{\partial v}\right)_{s,x}=0\ ; \qquad \text{(ii)}\ \left(\frac{\partial \mathrm{W}(v, s, x)}{\partial s}\right)_{v,x}=0. \quad . \quad . \quad (1)$$

In order to obtain the variation of c/a with x, as determined experimentally, we should have to eliminate v between these two equations. We shall make the assumption that the changes of volume arise from forces which act equally in all directions and can thus be simulated by the effect of a uniform pressure or tension. Such a uniform pressure or tension, in an anisotropic elastic crystal, produces a shearing strain given by

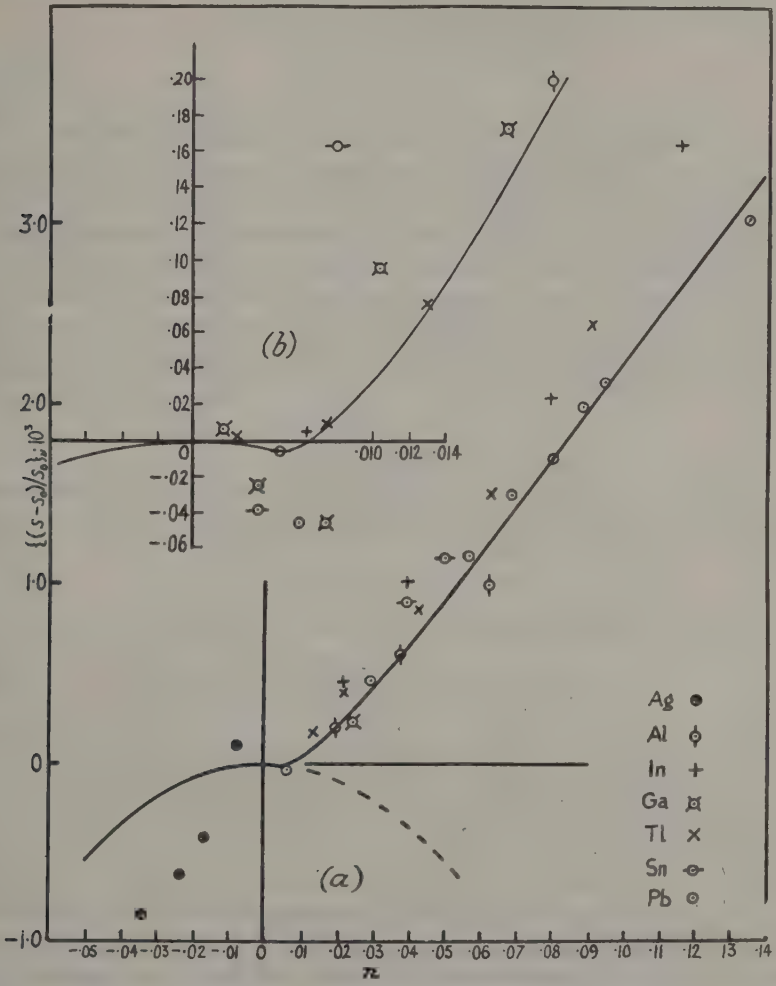
$$\frac{s-s_0}{s_0}=\tfrac{2}{3}k\left(\frac{v-v_0}{v_0}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where

$$k=\frac{s_{13}+s_{33}-s_{11}-s_{12}}{2(s_{11}+s_{12})+4s_{13}+s_{33}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and the s_{ij} are the elastic *coefficients* of the crystal. Thus, if we subtract this strain from the observed value, we get the strain induced by the solute atoms at *constant volume*. This quantity which we can determine from the experimental results can now be compared with the relation between s and x given by the second of equations (1) alone. The reason for this procedure is that the dependence of W on v is too complicated to be amenable to calculation, whilst the dependence of s on x at constant v, we believe, arises from a particular effect which can be comparatively easily isolated.

Fig. 1.



(*a*) The variation of the c/a ratio with electron concentration in Mg solid solutions.
(*b*) Detail of (*a*) for very small concentrations.

If $\delta a/a$ and $\delta c/c$ are the observed strains, we get, for the change of s at constant volume,

$$\left(\frac{s-s_0}{s_0}\right)_v = \tfrac{2}{3}\left(\frac{\delta c}{c}-\frac{\delta a}{a}\right)-\tfrac{2}{3}k\left(\frac{\delta c}{c}+2\frac{\delta a}{a}\right)$$

$$=\tfrac{2}{3}(1-k)\frac{\delta c}{c}-\tfrac{2}{3}(1+2k)\frac{\delta a}{a}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

For pure magnesium $k=0{\cdot}0226$. In fig. 1 $\{(s-s_0)/s_0\}_v$ is plotted, not against atomic concentration, but against the increase of the number of electrons per atom in the alloy which we denote by n. If $\mathcal{N}$ is the number of valence electrons per atom of the solute, $n=(\mathcal{N}-2)x$. The values of $\mathcal{N}$ assumed for the solute metals are as follows:

	Li	Ag	Al	Ga	In	Tl	Sn	Pb
$\mathcal{N}=$	1	1	3	3	3	3	4	4

§ 3. A THEORY OF THE ORIGIN OF THE SHEARING STRAIN.

Since fig. 1 shows clearly that variations of s at constant volume depend only on n, we shall use, from this point onwards, n as the variable in place of x; and also since all changes, now to be considered, take place at constant volume, there is no further need to make explicit reference to v, and therefore this quantity will no longer be shown as a variable. The condition for equilibrium becomes therefore $(\partial W(s, n)/\partial s)_n=0$, which is the required equation to determine s as a function of n. For the derivative of s with respect to n, we have therefore

$$\frac{ds}{dn}=-\frac{\partial^2 W}{\partial s \partial n}\Big/\left(\frac{\partial^2 W}{\partial s^2}\right)_n. \qquad (5)$$

Since (W/v) is equal to the strain energy, apart from a constant, and since the strain energy in terms of s can easily be shown to be $\frac{1}{2}\mu(s-s_0)^2/s_0^2$, where

$$\mu=\tfrac{1}{2}(c_{11}+c_{12}+2c_{33}-4c_{13}),$$

and c_{ij} are the elastic *constants* of the crystal, we find

$$\frac{1}{s_0}\frac{ds}{dn}=-\left(\frac{s_0}{\mu v}\right)\frac{\partial^2 W}{\partial n \partial s}. \qquad (6)$$

It will now be assumed that the only part of W which contributes to the second derivative on the right-hand side of (6) is the Fermi energy. This Fermi energy per atom, which we denote by w_F, may be considered as made up of several parts as follows:

$$w_F=w_0+\Sigma w_i-\Sigma w_j, \qquad (7)$$

where w_0 is the energy corresponding to the completely filled Brillouin zone of two electrons per atom. In magnesium or its alloys some electrons must overlap into the second zone, at various points, leaving vacancies in the first zone. The overlapping electrons and the holes may lie in distinct regions of momentum space. The energy of one such region containing n_i overlapping electrons is denoted by w_i and the energy of n_j electrons which would just fill one particular region of holes is written w_j. Thus we see that w_F is given by (7), where the summations are over all distinct regions of holes (j) and overlapping electrons (i).

w_i is a function of s and n_i only, but n_i is itself a function of s and n, and therefore

$$\left(\frac{\partial w_i}{\partial s}\right)_n=\left(\frac{\partial w_i}{\partial s}\right)_{n_i}+\left(\frac{\partial w_i}{\partial n_i}\right)_s\left(\frac{\partial n_i}{\partial s}\right)_n. \qquad (8)$$

Now $(\partial w_i/\partial n_i)_s = \mathrm{E}_i(s, n_i)$, where E_i is the energy at the top of the filled levels in the region (i). For equilibrium, we have

$$\mathrm{E}_1 = \mathrm{E}_2 = \ldots\ldots = \mathrm{E}_i = \ldots\ldots = \mathrm{E}_j = \zeta, \quad \ldots\ldots \quad (9)$$

where ζ is the energy at the surface of the filled states, and thus

$$\Sigma \mathrm{E}_i \left(\frac{\partial n_i}{\partial s}\right)_n - \Sigma \mathrm{E}_j \left(\frac{\partial n_j}{\partial s}\right)_n = 0, \quad \ldots\ldots \quad (10)$$

since

$$n = \Sigma n_i - \Sigma n_j. \quad \ldots\ldots \quad (11)$$

Thus, since w_0 is independent of n, we have, using (8) and (10).

$$\frac{\partial^2 w_{\mathrm{F}}}{\partial n \partial s} = \Sigma \frac{\partial^2 w_i}{\partial s \partial n_i} \left(\frac{\partial n_i}{\partial n}\right)_s - \Sigma \frac{\partial^2 w_j}{\partial s \partial n_j} \left(\frac{\partial n_j}{\partial n}\right)_s. \quad \ldots \quad (12)$$

If $\mathrm{N}(\zeta)$ denotes the total density of states at the energy ζ, and $\mathrm{N}_i(\zeta)$ the contribution from the region of momentum space (i), then clearly

$$\left(\frac{\partial n_i}{\partial n}\right)_s = \frac{\mathrm{N}_i(\zeta)}{\mathrm{N}(\zeta)}, \quad \ldots\ldots \quad (13)$$

with exactly similar relations, apart from a negative sign, holding for the regions of the holes.

Thus (12) becomes

$$\frac{\partial^2 w_{\mathrm{F}}}{\partial n \partial s} = \frac{1}{\mathrm{N}(\zeta)} \left\{ \Sigma \mathrm{N}_i \left(\frac{\partial \mathrm{E}_i}{\partial s}\right)_{n_i} + \Sigma \mathrm{N}_j \left(\frac{\partial \mathrm{E}_j}{\partial s}\right)_{n_j} \right\}. \quad \ldots \quad (14)$$

Finally, if we define coefficients ϕ_i by the equations

$$\phi_i = \frac{s_0}{\zeta} \left(\frac{\partial \mathrm{E}_i}{\partial s}\right)_{n_i}, \quad \ldots\ldots \quad (15)$$

(6) and (14) give

$$\frac{1}{s_0} \frac{ds}{dn} = -\left(\frac{\zeta}{v\mu}\right) \frac{1}{\mathrm{N}(\zeta)} \{\Sigma \phi_i \mathrm{N}_i + \Sigma \phi_j \mathrm{N}_j\}. \quad \ldots \quad (16)$$

§ 4. COMPARISON WITH EXPERIMENTAL RESULTS.

In order to apply equation (16) to the magnesium solid solutions, we have to make detailed assumptions with regard to where the holes and the overlapping electrons occur in relation to the Brillouin zone. Fig. 2 (*a*) shows the form of the zone which contains, for all values of s, exactly two electrons per atom. We shall assume that the holes in the first zone occur at points along the line $\mathrm{D}_1\mathrm{D}_2$ and regard the centre of this line D as the point of highest energy *. We shall also assume that electrons overlap at the points A, and that for $n > 0$ overlap may also occur at the points B.

All points like those indicated by D in fig. 2 (*a*) are equivalent and lie at equal distances from the origin for all values of s. The neighbourhoods of D form altogether six identical regions in which the surfaces of constant energy are simply connected closed surfaces. There are two separate regions corresponding to the points A, and one to the points B. The distances from the origin, k_A, k_B and k_D, are given by :

---

* Until calculations of the regions of highest energy within the Brillouin zone have been made any assumption such as this is necessarily arbitrary since the geometry of the zone is not sufficient to indicate the relative energies. It is of some interest that the coefficient ϕ_{D} depends upon the assumed position of the point of highest energy and that the experimental data relating to the variation of c/a are most conveniently explained by the present choice.

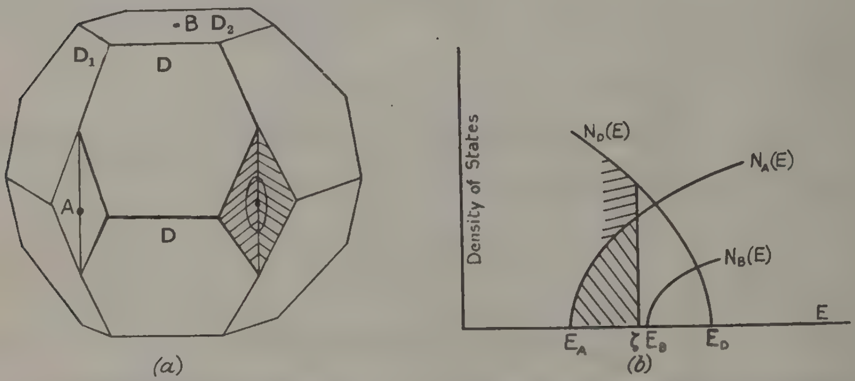
$$k_A=(\sqrt{3}/4v)^{1/3}\tfrac{2}{3}s^{1/2},$$

$$k_B=(\sqrt{3}/4v)^{1/3}\frac{1}{s},$$

$$k_D=(\sqrt{3}/4v)^{1/3}\frac{1}{\sqrt{3}}\left(1+\frac{3}{4s^3}\right)s^{1/2}.$$

In order to calculate the coefficients ϕ_i, we notice, first, that $\partial E_i/\partial s$ is taken at constant n_i and, therefore, the changes in E_i must be closely equal to the changes in E_i when $n_i=0$, *i. e.* of the energy of the centre point of the region, and secondly, in equilibrium the value of E_i is always equal to ζ.

Fig. 2.



(*a*) Brillouin zone for Mg.

(*b*) Density of states for overlapping electrons N_A and for holes N_D.

For these reasons, we may assume that ϕ_A, for example, will be given without serious error by $d\log E_A/d\log s$. It is to be noticed that the ϕ's are unlikely to depend at all sensitively on n, and, of course, according to the above assumption, they are completely independent of n. For the application of (16), this is their most important property since it means that the right-hand side of this equation only varies with n significantly through the densities of states N_i. Numerical estimates of the ϕ's may be obtained by assuming that E_A, E_B and E_D vary as the squares of the distances from the origin to the points A, B and D respectively. In this way we find :

$$\phi_A=1, \qquad \phi_B=-2, \qquad \phi_D=1-\frac{18}{3+4s^3}=-\frac{13}{41}.$$

Thus (16) may be written

$$\frac{1}{s_0}\frac{ds}{dn}=\left(\frac{\zeta}{v\mu}\right)\frac{1}{N}\left\{2N_B+\frac{78}{41}N_D-2N_A\right\}, \quad . \quad . \quad . \quad (17)$$

where N_A, N_B and N_D denote the densities of states of *single* regions about their respective points.

Since, even for $n=0$, there is overlap at A and holes at D, we can expand N_A and N_D about their values at $n=0$. Thus, for instance,

$$N_A(n)=N_A^{(0)}+\left(\frac{dN_A}{dE}\right)_0\left(\frac{dE}{dn}\right)_0 n$$

$$=N_A^{(0)}+\frac{1}{N}\frac{dN_A}{d\zeta}n, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where N is the total density of states in pure magnesium. N_B cannot, of course, be expanded in this way, since it is zero up to a certain (very small) value of n which we denote by n_0. For $n>n_0$, we may assume that

$$N_B=\alpha\sqrt{(\zeta-E_B)},$$

where α is a constant. Also we have $n-n_0=(\zeta-E_B)N$, and therefore

$$N_B=\frac{\alpha}{\sqrt{N}}(n-n_0)^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Finally, substituting for N_A, N_B and N_D into equation (17), and integrating with respect to n, we find

$$\left(\frac{s-s_0}{s_0}\right)_v=A(n-n_0)^{3/2}+Bn+Cn^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

where

$$A=\begin{cases}\frac{4}{3}\left(\frac{\zeta}{\mu v}\right)\frac{\alpha}{N^{3/2}}, & \text{if } n>n_0,\\ 0\quad, & \text{if } n<n_0,\end{cases} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$B=\frac{2}{N}\left(\frac{39}{41}N_D^{(0)}-N_A^{(0)}\right)\left(\frac{\zeta}{\mu v}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

and

$$C=\frac{1}{N^2}\left(\frac{39}{41}\frac{dN_D}{d\zeta}-\frac{dN_A}{d\zeta}\right)\left(\frac{\zeta}{\mu v}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

The experimental results, for solutes with $\mathcal{N}>2$, indicate (*cf.* fig. 1) that B must be zero or exceedingly small. Hence we deduce that $N_A^{(0)}$ and $N_D^{(0)}$ must be very nearly equal.

It is clear from equation (23) that C must be negative, for $dN_D/d\zeta$ is necessarily negative and $dN_A/d\zeta$ positive (see *e.g.* fig. 2 (*b*)). Thus monovalent solutes, for which $n<0$, must decrease the c/a ratio. This conclusion is in agreement with the experimental results for lithium and silver.

It is possible to make a direct estimate of the number C by assuming that N_A and N_D have approximately parabolic form as shown in fig. 2 (*b*). If ϵ_1 is the energy width of the electrons overlapping at A, and ϵ_2 the energy width of the distribution of holes, then

$$\frac{dN_A}{d\zeta}=\frac{N_A}{2\epsilon_1}, \qquad \frac{dN_D}{d\zeta}=-\frac{N_D}{2\epsilon_2}.$$

For magnesium $\zeta = 7{\cdot}3$ eV. as given by the soft X-ray emission spectra (Skinner 1939), $\mu = 6{\cdot}23 \times 10^{11}$ dynes/cm.^{2}, and $v = 23{\cdot}08 \times 10^{-24}$ cm.^{3} Thus the pure number $(\zeta/\mu v) = 0{\cdot}813$. Hence, neglecting the difference between 39/41 and unity, we obtain

$$C = -\frac{0{\cdot}813}{2N^2}\left(\frac{N_A}{\epsilon_1} + \frac{N_D}{\epsilon_2}\right),$$

and, since $N = 6N_D + 2N_A$, and $N_A = N_D$, we get, if we assume that N may be taken as $3/\zeta$, which is the free electron value and probably slightly too big,

$$C = -\frac{0{\cdot}813\zeta}{48}\left(\frac{1}{\epsilon_1} + \frac{1}{\epsilon_2}\right).$$

If we take $\epsilon_1 = \epsilon_2 = 1$ eV. which is roughly the value of ϵ_1 suggested by the X-ray data, we find $C = -0{\cdot}247$. The value of C which has been used to plot the curves of fig. 1 is $-0{\cdot}213$, and is thus quite compatible with the above estimate. The curves of fig. 1 are given by

$$\begin{aligned}\left(\frac{s - s_0}{s_0}\right)_v &= -0{\cdot}213n^2 + 0{\cdot}150(n - 0{\cdot}005)^{3/2}, \qquad && n > 0{\cdot}005,\\ &= -0{\cdot}213n^2 \qquad\qquad\qquad\qquad\qquad , && n < 0{\cdot}005,\end{aligned}$$

where it has been assumed that overlap sets in at the points B when n exceeds 0·5 per cent.

The agreement is not good in the case of silver, but the curve for negative n goes exactly through the only known point for the magnesium lithium solid solution, viz. that at the phase boundary limit of 17 atomic per cent lithium *. At this concentration the observed $(s - s_0)/s_0$ is equal to $-{\cdot}0067$.

It must be admitted that the interpretation of the experimental results, which we have adopted, depends largely on the observations at very small n. Apart from this region, a fairly good straight line could be drawn through the whole set of points both for positive and negative n. The interpretation would then be that the constants A and C were relatively small, and that B was positive, implying that $N_D > N_A$. To distinguish with certainty between these possibilities more data relating to the strain produced by monovalent solutes would be necessary.

REFERENCES.

HUME-ROTHERY, W., and RAYNOR, G. V., 1940, *Proc. Roy. Soc.*, **177,** 27.
JONES, H., 1949, *Physica*, **15,** 13.
RAYNOR, G. V., 1940, *Proc. Roy. Soc.*, **174,** 457 ; 1942, *Ibid.*, **180,** 107.
SKINN H. W. B., 1939, *Rep. Prog. Phys.*, **5,** 257 (London: Physical Society).

---

* The author is indebted to Professor G. V. Raynor for the information relating to the Mg-Li system.

LX. *On the Correlation of the Directional Properties of Rolled Shee in Tension and Cupping Tests.*

By L. Bourne, B.Met., and R. Hill, M.A., Ph.D.,
Metal Flow Research Laboratory, Sheffield*.

[Received April 12, 1950.]

Summary

Plastic anisotropy in rolled sheet is examined by tension and cupping tests. The earing positions are correlated with the strain-ratios measured in tension tests at various orientations to the direction of rolling. The materials used are copper giving four ears at 45°, brass giving four ears at 50°, and brass giving six ears at 0° and 60°. A theory of plastic anisotropy due to Hill is found to be in good agreement with the experimental data for materials producing four ears. The theory is extended to describe more complex states of anisotropy.

§ 1. Method of Correlation.

The deformation imparted during the rolling of sheet metal produces an elongation of the crystal grains and a tendency towards a preferred orientation. The resulting anisotropy can be reduced or accentuated at will by a suitable programme of rolling reductions and interstage heat-treatment. Directionality in the final product is revealed by earing in a cupping test on a circular blank, or by the variation of mechanical properties in tension tests on specimens cut from the sheet at different angles to the direction of rolling. It is evident that the plastic behaviour of a metal with crystallographic anisotropy is directly dependent on the plastic properties of a single crystal of the metal and on the degree of preferred orientation. If the mechanism of distortion of a grain in an aggregate were completely known, it would be possible, in principle, to predict the behaviour of the aggregate when loaded in any way. This cannot yet be done since it is not known, for instance, how the operative glide system in an individual grain is influenced by the constraint of neighbouring grains. For example, in a face-centred cubic metal there are twelve operable slip-directions at ordinary temperatures, whereas a general uniform strain can be produced by a combination of shears in only five independent directions. Sufficient conditions for the determination of a unique active system are unknown. For this reason alone, preliminary attempts to relate the earing positions to the rolling texture seem to be premature†.

---

* Communicated by the Authors.

† For a general review of past work see T. Ll. Richards, " Progress in Metal Physics ", 1 (1949), pp. 281–305. (Butterworth's Scientific Publications, London.)

A more practical alternative is to measure the macroscopic properties of the anisotropic sheet in simple tests and, from the data so obtained, attempt to predict the behaviour of the sheet when deformed by any complex stress system. The question immediately arises as to which properties of the sheet are relevant to its behaviour under combined stress, in particular when deep-drawn into a cup. It is obvious that only *mechanical* properties are relevant, and that of these the breaking stress and elongation can be ignored when we are not concerned with fracture. Again, when we are interested only in *large* plastic deformations, the elastic moduli are irrelevant except in so far as they affect the residual stresses after unloading. There remains only the plastic stress–strain properties, which incorporate (i) the dependence of work-hardening on the amount of deformation, and (ii) the relations between the ratios of the components of the stress and the plastic strain-increment. In general, both factors govern the distortion of the sheet when taken along some complex loading-path. In the cupping test, however, the conditions are fairly simple, each element of the rim being subjected to a circumferential strain of approximately the same amount if the material is not too anisotropic. Moreover, ear-formation presumably depends mainly (though not entirely) on conditions near the rim, and here the state of stress is essentially a pure circumferential compression. The blank-holder pressure is usually negligible compared with the yield stress, and hinders the thickening of the edge only slightly because of the elasticity of the apparatus. Hence factor (ii) appears likely to predominate in controlling the positions of the ears. The relevant fundamental test is therefore the measurement of strain-ratios in uniaxial compression at various angles in the plane of the sheet. For convenience it is obviously preferable to perform tension tests instead, although this involves the assumption that the strain-ratios are unaffected by microscopic internal stresses which exert different influences in tension and compression. This seems entirely reasonable since the final annealing should remove those stresses responsible for the Bauschinger effect and allied phenomena. The measurement of strain-ratios in tension tests is the basis of the present method of predicting earing positions. In essence, the method has been used previously by Baldwin, Howald and Ross (1945), but the details differ, as will be explained later.

§ 2. THEORETICAL CONSIDERATIONS.

According to a theory of plastic anisotropy proposed by Hill (1948), the relations between the ratios of the components of the stress and strain-increment are

$$\frac{d\epsilon_x}{(G+H)\sigma_x-H\sigma_y}=\frac{d\epsilon_y}{(F+H)\sigma_y-H\sigma_x}=\frac{d\gamma_{xy}}{N\tau_{xy}}, \quad . \quad . \quad . \quad (1)$$

where the x-axis is the direction of rolling and the y-axis is transverse to this in the plane of the sheet. The applied stress is considered to be in the plane of the sheet, with components σ_x, σ_y, and τ_{xy}. F, G, H, and

N are parameters specifying the current state of anisotropy, which will normally vary during a process of plastic distortion. Similar linear relations have been independently derived, though from different physical assumptions, by Jackson, Smith and Lankford (1949) and by Dorn (1949). Dorn's relations contain five parameters, and cannot be derived from a plastic potential (see equation (10) below).

(i) *Strain-ratios in a tension test.*—For a uniaxial tension σ applied at an angle α to the direction of rolling, the (x, y) components of stress are

$$\sigma_x=\sigma \cos^2 \alpha, \qquad \sigma_y=\sigma \sin^2 \alpha, \qquad \tau_{xy}=\sigma \sin \alpha \cos \alpha. \quad . \quad . \quad (2)$$

The strain-increment in the width direction of the tension specimen, with current width w, is

$$\frac{dw}{w}=d\epsilon_x \sin^2 \alpha+d\epsilon_y \cos^2 \alpha-2d\gamma_{xy} \sin \alpha \cos \alpha, \quad . \quad . \quad . \quad (3)$$

while the strain-increment through the thickness t is

$$\frac{dt}{t}=-(d\epsilon_x+d\epsilon_y) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

since there is no change of volume. Combining these four equations we obtain

$$\frac{dw/w}{dt/t}=\frac{\mathrm{H}+(2\mathrm{N}-\mathrm{F}-\mathrm{G}-4\mathrm{H}) \sin^2 \alpha \cos^2 \alpha}{\mathrm{F} \sin^2 \alpha+\mathrm{G} \cos^2 \alpha}. \quad . \quad . \quad . \quad (5)$$

If the state of anisotropy is not altered by the deformation during the test, the ratios of the parameters remain constant and (5) can be integrated to give

$$\frac{\ln (w_0/w)}{\ln (t_0/t)}=\frac{\mathrm{H}+(2\mathrm{N}-\mathrm{F}-\mathrm{G}-4\mathrm{H}) \sin^2 \alpha \cos^2 \alpha}{\mathrm{F} \sin^2 \alpha+\mathrm{G} \cos^2 \alpha}=r(\alpha), \text{ say}, \quad . \quad (6)$$

where w_0 and t_0 are the initial width and thickness. The ratio $r(\alpha)$ between the logarithmic strains in the width and thickness directions then has a constant value throughout any one test. In a completely isotropic material $\mathrm{F}=\mathrm{G}=\mathrm{H}=\mathrm{N}/3$ and $r(\alpha)\equiv 1$. If the sheet is isotropic only in its plane, $\mathrm{N}=\mathrm{F}+2\mathrm{H}=\mathrm{G}+2\mathrm{H}$, $r(\alpha)\equiv \mathrm{H}/\mathrm{F}$. In a non-isotropic material the ratios H/G and H/F can be found from tension tests along and perpendicular to the direction of rolling. Thus :

$$\frac{\mathrm{H}}{\mathrm{G}}=r(0), \qquad \frac{\mathrm{H}}{\mathrm{F}}=r(\pi/2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

A third ratio, N/H say, can be obtained from a test in some other direction. The theoretical relation (6) is then completely known, and its validity can be checked by comparing it with the experimental results over the whole range of orientations.

Some general properties of (6) are worth noting. Stationary values of $r(\alpha)$ occur when $\alpha=0°$, 90°, and also for a certain intermediate value β except when $(2\mathrm{N}-\mathrm{F}-\mathrm{G}-4\mathrm{H})/\mathrm{H}$ lies between $(\mathrm{F}-\mathrm{G})/\mathrm{G}$ and $(\mathrm{G}-\mathrm{F})/\mathrm{F}$.

The intermediate stationary value is a maximum when $(2N-F-G-4H)/H$ is positive, and a minimum when it is negative. In particular, when the properties are the same in both the rolling and transverse directions, so that $F=G$, the $r(\alpha)$ relation is symmetrical about $\beta=45°$; its value at the latter is a maximum if $N>F+2H$ and a minimum if $N<F+2H$.

In an anisotropic specimen a tensile stress generally produces a shear strain as well as an extension; that is, a principal strain axis does not usually coincide with the direction of the applied stress. It is easily shown from the above equations that the shear-strain increment, referred to axes along and perpendicular to the applied stress, is equal to

$$\frac{dw}{w}\times\frac{[(N-F-2H)\sin^2\alpha-(N-G-2H)\cos^2\alpha]\sin\alpha\cos\alpha}{[H+(2N-F-G-4H)\sin^2\alpha\cos^2\alpha]}\quad . \quad . \quad (8)$$

The presence of a shear strain has been demonstrated experimentally by Hazlett, Robinson and Dorn (1949). In the present tests the angle of shear, as computed from (8) with the deduced values of the parameters, was never more than about 5°, even at the largest extensions. The corresponding eccentricity of loading was negligible in view of other experimental errors and the unavoidable scatter in the results. It can be seen from (8) that the axes of stress and strain-increment coincide when $\alpha=0°$, $90°$, or an intermediate value γ satisfying

$$\tan^2\gamma=\frac{N-G-2H}{N-F-2H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

γ coincides with the stationary point β of the $r(\alpha)$ relation only when $F=G$.

(ii) *Angles of necking.*—A *sufficiently thin* flat strip necks at an oblique angle when pulled in tension. In a perfectly isotropic specimen the angle is about 55° to the axis. According to theory (Hill 1948), the line of the neck develops in the direction of zero rate of extension in the uniformly deforming strip. In a non-isotropic specimen the two directions of zero extension-rate make different angles with the axis, except when $\alpha=0°$, $90°$, or γ. If, then, the neck does in fact coincide with a direction of zero extension, a further check on the preceding theory is provided by measurements of the necking angles. For example, for a test in the rolling direction the theoretical inclination of the neck to the axis is $\tan^{-1}\sqrt{(1+G/H)}$.

Unfortunately the fractures in the copper and brass specimens used in the present experiments were too ragged for sufficiently precise measurements of the necking-angles. It appeared, however, that the necks usually formed at greater angles to the specimen axis than would be expected theoretically. This discrepancy is probably due to a slight necking in the width direction, which occurred towards the end of the considerable range of uniform strain and extended over a large part of the gauge length. The theory seems to give closer agreement in a metal where the range of uniform strain is comparatively small (Hill 1948).

(iii) *Earing positions.*—The state of stress and strain in a circular blank which is being deep-drawn into a cup can, in principle, be calculated from equations (1) along with some assumed law of work-hardening under combined stress. The positions and gradual development of the ears would be included in the results of the analysis. However, although the calculation can be made for an isotropic blank (Hill 1949) the analysis becomes extremely involved when the blank is anisotropic. Consequently, some simpler method must be adopted for predicting the earing positions.

In view of the approximate symmetry of the deformation about the tip of an ear or the bottom of a hollow, it seems a reasonable assumption (Hill 1948) that the ears and hollows develop at the points on the rim where the principal axes of the stress and the strain-increment coincide. Since the stress in the rim is a circumferential compression, it follows from the previous theory that the ears and hollows should begin to form at points on the rim whose angular positions are 0°, 90° and $90°-\gamma$, relative to the direction of rolling. There are therefore four ears and four hollows (at most), and the equations (1) must be modified to represent the anisotropy of a material giving more than four ears (see below). Now, when the yield criterion and stress–strain equations are connected on the plastic potential basis, Hill has shown that the initial yield stress has stationary values in the directions for which the axes of stress and strain coincide. It is natural to assume, therefore, that the ears form in the positions where the (circumferential)yield stress is initially a minimum (numerically) and the hollows where it is a maximum. Then, if N is greater than both $F+2H$ and $G+2H$ the ears are in the $90°-\gamma$ positions, and if N is less than both $F+2H$ and $G+2H$ the ears are in the 0° and 90° positions. If N is intermediate to $F+2H$ and $G+2H$ only two ears are formed; these are in the 0° positions if $F>G$ and in the 90° positions if $F<G$.

In particular, if $F=G$, there are four ears in the 45° positions if $N>F+2H$, and in the 0° and 90° positions if $N<F+2H$. These are also the positions where the strain-ratio r is a maximum in the special case $F=G$. This, indeed, is what would be expected if the circumferential strain in the blank did not depend on the angular coordinate, since the ears correspond to the parts of the blank which suffer most radial extension. When applied generally, this consideration leads to the hypothesis that the ears develop at positions where the strain-ratio r is a maximum for a uniaxial stress in the *circumferential* direction. However, this alternative hypothesis is unlikely to be correct since in actuality the circumferential strain varies considerably around the rim of a non-isotropic blank. In fact, the experimental evidence reviewed later is more in accord with the assumption that the ears form at positions where the strain-ratio r is a maximum for a uniaxial stress in the *radial* direction. This third hypothesis was suggested on empirical grounds by Baldwin Howald and Ross, but it is difficult to see a reasonable theoretical basis for it.

It is worth remarking that the absence of ears in a cupping test does not prove that the sheet is completely isotropic, but only isotropic in its plane. According to the above theory this implies $N=F+2H=G+2H$, but H is not necessarily equal to F and G. To prove complete isotropy it is essential to perform, in addition, a tensile test in the plane of the sheet to see whether the strain-ratio is unity or not.

With suitable rolling and heat treatment some metals can be made to give six ears in a cupping test, *e. g.* brass, or even eight ears, *e. g.* certain aluminium alloys (Chevigny 1949). To represent these more complex states of anistropy it is necessary to generalize the stress–strain equations (1). This may be achieved (Hill 1950) by writing

$$\frac{d\epsilon_x}{\partial f/\partial\sigma_x}=\frac{d\epsilon_y}{\partial f/\partial\sigma_y}=\frac{d\gamma_{xy}}{\partial f/\partial\tau_{xy}}, \quad \ldots\ldots\ldots \quad (10)$$

where f, the plastic potential, is taken to be a polynomial of degree n in the components of stress*. f is homogeneous if the strain-ratios do not depend on the absolute magnitude of the stress. The corresponding number of ears is $2n$ (or less, in special circumstances). The plastic potential for equations (1) is the quadratic

$$(G+H)\sigma_x^2-2H\sigma_x\sigma_y+(H+F)\sigma_y^2+2N\tau_{xy}^2. \quad \ldots\ldots \quad (11)$$

To represent a material giving six ears, f must be a cubic (at least) and for a material giving eight ears f must be a quartic (at least).

§ 3. MATERIALS AND TREATMENTS.

The materials used were copper and brass. In pure copper there are two common recrystallization textures, viz. (100) [001] giving four ears at 0° and 90°, and (113) [$\bar{2}\bar{1}$1] giving four ears at 45°. Only the latter was studied here since the former has been thoroughly investigated by Baldwin, Howald and Ross. In 70/30 brass the usual earing positions are at 45° or at 0° and 60°. Both these were examined.

The copper was phosphorus deoxidized, with an average grain size of 0·06 mm. It was received as a hot-rolled and pickled strip of width 7 in. and thickness $\frac{3}{8}$ in. The strip was machined to 0·210 in. to remove surface pitting, and rolled to a nominal thickness of 0·030 in. (85 per cent reduction). It was then annealed at 650° C. for half an hour.

The brass was received as a hot-rolled strip of width 7 in. and thickness $\frac{3}{8}$ in. To obtain a texture giving four ears the strip was rolled 50 per cent, annealed at 425° C., rolled 84 per cent, and annealed at 650° C. The final nominal thickness was 0·030 in. To obtain a texture giving six ears the strip was rolled 50 per cent, annealed at 650° C., rolled 84 per cent, and annealed at 650° C. The final thickness was again 0·030 in. Burghoff and Bohlen (1942) have determined the (111) pole figures for both treatments.

---

* In the differentiation τ_{xy} and τ_{yx} are treated as distinct, the terms containing the shear–stress component being regarded as split into two equal parts. Only even powers of the shear–stress are allowable since the axes of anisotropy are the axes of reference.

In the cupping tests 4 in. diameter blanks were drawn with a punch of diameter 2 in. over a die of radius 0·187 in. The nominal blank-holder clearance was 0·003 in. The lubricant was a mixture of tallow and graphite. Cups drawn from the copper strip had four large ears at 45° to the rolling direction. The four-eared brass cups had large ears at approximately 50° to the direction of rolling, and the six-eared brass cups had small ears at 0° and 58° ($\pm 2°$).

The tensile test pieces were produced in a blanking press and milled to size, the edges being polished on emery paper. Specimens (in duplicate) were cut at 0°, $22\frac{1}{2}°$, 45°, $67\frac{1}{2}°$ and 90°. The gauge length was 2 in. and the width 0·5 in. The tests were carried out in a Hounsfield tensometer. Extensions in the elastic and initial plastic range were measured by a Lindley extensometer. Width and thickness measurements (average of three along the gauge length) were taken with a hand-micrometer reading to 0·0001 in. The readings were continued up to the stage at which necking began.

§ 4. Experimental Results.

With all the materials tested it was found that the strain-ratio r did not vary with strain, within experimental scatter. The constancy of r was maintained until necking intervened ; this occurred at extensions of the order of 30–40 per cent. The inference to be drawn is that the initial anisotropy in the rolled strip was not appreciably altered by the uniform tensile strains.

Values of $r(\alpha)$, derived from the slopes of the best straight lines through the experimental points (means of at least two tests), are shown in fig. 1. Variations in the material, errors of measurement, and scatter in the results, combine to produce an uncertainty of order $\pm 0{\cdot}1$ in the derived values of r. Ratios of F, G, H, N were obtained by fitting (6) to the results for the copper and brass giving four ears. There is some uncertainty in the values of the parameters, amounting perhaps to $\pm 0{\cdot}1$, owing to the arbitrariness in choosing the best fit.

The curve in fig. 1 (*a*) for the copper corresponds to $F=G=H=N/4$. The theory in § 2 (iii) predicts four ears at 45° when $F=G$ and $N>F+2H$, in agreement with experiment. Comparison may be made with the data of Baldwin, Howald and Ross, for copper giving 45° ears (see the lower curve in fig. 18 of their paper). The parameters for their material, which had received a different treatment, are $F=G=H=N/5$.

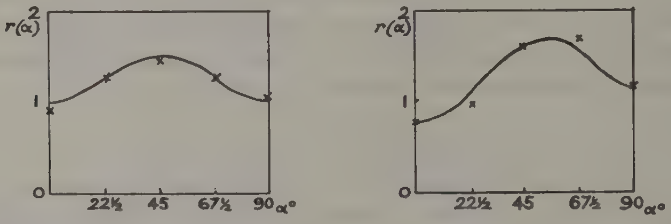
The parameters for the brass giving four ears, corresponding to the curve in fig. 1 (*b*), are

$$F:G:H:N=1:1{\cdot}6:1{\cdot}2:5{\cdot}5.$$

The angle γ in (9) then has the value 40°, approximately, so that $90°-\gamma=50°$. Since N is greater than both $F+2H$ and $G+2H$, the theory predicts four ears at 50°. This is what was observed. It is, of course, hardly possible to be precise about the actual angular position of the tip of an ear, especially since there are always slight differences between each of the four ears. In view of the uncertainty in the derived values of the parameters, a discrepancy of less than 5° between theory and experiment would probably not be significant.

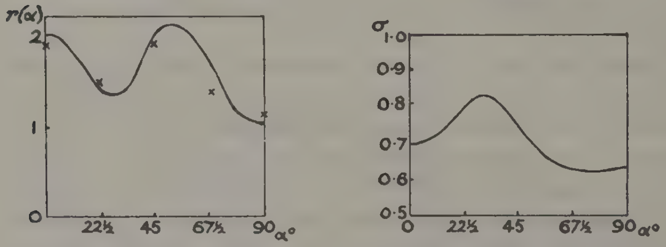
As a further check on the validity of equation (1), a compression test normal to the sheet was carried out on a cylindrical brass disc of diameter 0·25 in. and thickness 0·029 in. The section became elliptical, with the major axis in the direction of rolling. After a 50 per cent compression the axes were 0·38 in. and 0·325 in. (approximately, since the edge became ragged)*. The corresponding ratio of the logarithmic strains is 1·6. Regarding the test as equivalent to the application of equal biaxial tensions in the plane of the sheet, equation (1) gives $d\epsilon_x/d\epsilon_y = G/F = 1{\cdot}6$, in perfect agreement.

Fig. 1.



(*a*) Copper with four ears at 45°.

(*b*) Brass with four ears at 50°



(*c*) Brass with six ears at 0° and 58°.

(*d*) Theoretical variation of yield stress with orientation in brass with six ears.

In fig. 1 (*c*), for the brass giving six ears at 0° and 58°, the curve corresponds to the relation

$$r(\alpha) = \frac{6\cos^6\alpha + 6\sin^6\alpha + 38\sin^4\alpha\cos^2\alpha - 10\sin^2\alpha\cos^4\alpha}{3\cos^4\alpha + 6\sin^4\alpha + \sin^2\alpha\cos^2\alpha}.$$

This was obtained by assuming a homogeneous cubic polynomial for the plastic potential defined in equation (10), and choosing the ratios of the six numerical coefficients so that the following five conditions were satisfied: (i) $r=2$ in the rolling direction; (ii) $r=2$ in the 45° direction; (iii) $r=1$ in the transverse direction; (iv) the axes of stress and strain to coincide in a tension test at 30° to the direction of rolling; (v) the ratio of the strain in the rolling direction to the transverse strain to be 1·5 in a compression test normal to the sheet. The latter is regarded as equivalent to equal biaxial tension in the plane of the sheet. The value 1·5 is near to that observed in a compression test on a cylindrical disc of diameter 0·25 in. and height 0·03 in., lubricated with calcium oleate. After a 50 per cent compression the axes of the ellipse were 0·375 in. and 0·325 in., approximately.

---

* Calcium oleate was the lubricant (coefficient of friction ~0·015). With these specimen dimensions the ellipticity would not be significantly influenced by friction.

The plastic potential so determined has the form

$$f \equiv 3\sigma_x^3 - 6\sigma_x^2\sigma_y - 6\sigma_x\sigma_y^2 + 4\sigma_y^3 + (4\sigma_x + 21\sigma_y)\tau_{xy}^2. \quad . \quad . \quad . \quad (12)$$

According to the hypothesis of §2 (iii) the fourth condition ensures that either an ear or a hollow occupies the 60° position. The other possible positions of ears and hollows are calculated to be 0°, 90°, and $\cot^{-1}\sqrt{11}=17°$ approximately. It will be seen that the calculated curve passes close to the two experimental points not deliberately fitted, viz. at $22\frac{1}{2}°$ and $67\frac{1}{2}°$. This method of comparison was preferred to the obvious one of fitting the curve to all five points, and then comparing predicted and observed earing positions, because there is considerable latitude in fitting a cubic curve to so few points when they are uncertain to $\pm 0{\cdot}1$.

It remains to distinguish between the calculated positions of the ears and hollows. If the yield criterion appropriate to (12) is $f=$constant (depending on the strain-history), the ears are predicted at 0° and 60° if it is supposed that they form at points where the circumferential yield stress is a (relative) *maximum*. This is opposite to what was found above for copper and brass sheet giving four ears, where the ears form at points where the circumferential yield stress is, theoretically, an (absolute) minimum. We are unable to suggest an explanation of the difference; it may be an accidental product of the theory, depending in a complex way on the entire distribution of the stresses and strains in the initial stages of the cupping test. The calculated variation of the yield stress with the orientation of the tension specimen is shown in fig. 1 (*d*).

It is interesting to notice that the earing positions are close to the angles for which the strain-ratio r has maximum values for a *radial* tension. This is the hypothesis of Baldwin, Howald and Ross, referred to previously. The present experiments do not indicate whether this is the real factor determining earing, or whether it is merely that the maximum values of r happen to fall rather close to the actual earing positions on account of some more fundamental property (which may, indeed, be peculiar to brass and copper).

The variation of the strain-ratio r with orientation has been measured for an aluminium alloy (61S–T6) by Hazlett, Robinson and Dorn (1949). The composition was 0·25 per cent Cu, 0·6 per cent Si, 1·0 per cent Mg, and 0·25 per cent Cr. The alloy was solution heat-treated, quenched, and aged. Values of the parameters giving best fit of equation (6) to their data are F : G : H : N=1 : 1·15 : 0·65 : 2·25. Since N is only slightly different from F+2H or G+2H the shear strains (8) should be negligible, and this was found to be so. The earing positions were not measured. The same authors found that the strain-ratios for a magnesium alloy with only slight anisotropy varied during a tension test. Measurements of strain-ratios in an aluminium alloy, for tests both in and out of the plane of the plate, have been reported by Klingler and Sachs (1948).

Strain-ratios for low carbon steel have been measured by Lankford, Snyder and Bauscher (1950), who make the interesting suggestion that plastic anisotropy may sometimes be advantageous to good performance in an *unsymmetrical* pressing operation.

§ 5. STRESS–STRAIN CURVES.

For the correlation of stress–strain curves, obtained in tension tests at various angles to the direction of rolling, it is necessary to relate some measure of the amount of work-hardening to the previous strain-history. Assuming that the amount of hardening can be specified by a scalar quantity (which appears probable when the state of anisotropy does not vary), the natural measure to take is the value of the expression f, representing the yield criterion and plastic potential. Jackson, Smith and Lankford (1949) claim to have found evidence that the degree of hardening in anisotropic sheet steel is a function only of the amount of plastic work per unit volume (a hypothesis which is a reasonable approximation for some isotropic metals.) The present tests do not support this hypothesis in the cases of copper and brass.

Thus, the stress–strain curves at 0°, $22\frac{1}{2}$°, 45°, $67\frac{1}{2}$°, and 90° for the brass giving four ears did not vary from a mean by more than 0·8 tn/in.^{2} at any strain up to 50 per cent extension. In view of the variation between nominally similar specimens, there was scarcely any significant difference between the stress–strain curves in the various directions. Hence, if the hardening depended only on the amount of plastic work, the tensile yield stress should not vary significantly with direction in a given mechanical state. This clearly follows from the hypothesis that the same mechanical state can be produced by performing the same amount of work in tension tests in different directions; since the stress–strain curves coincide, the final yield stress is the same in all directions. This conclusion, however, is not consistent with the assumption that the yield criterion is f=constant, since this implies variations of the yield stress with direction. For example, according to (11), the ratio of the yield stresses in the 0° and 90° directions should be $\sqrt{\{(F+H)/(G+H)\}}$ or about 0·9; the ratio of the yield stresses in the 40° and 90° directions is 0·8, approximately.

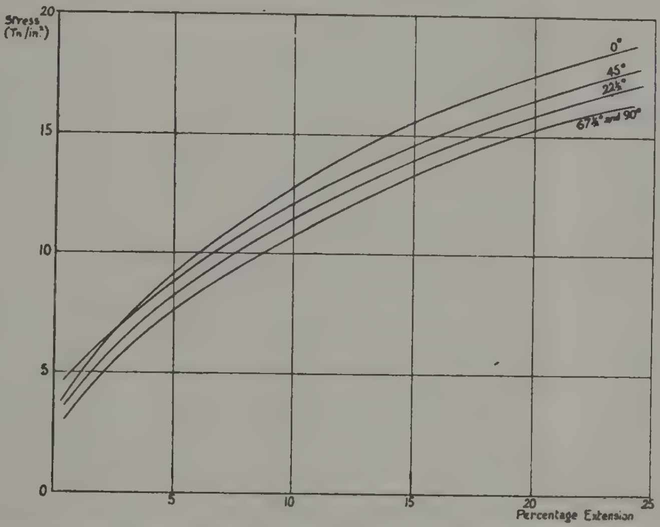
A similar conclusion applies also to the brass earing at 0° and 60°, where the stress–strain curves showed even less variation and the theoretical yield stresses more.

The stress–strain curves for the copper, however, showed significant differences (fig. 2), especially between the 0° and 90° curves. Since the theoretical yield stresses are roughly equal at 0° and 90°, the same conclusion is again reached, namely that the hardening (as measured by the value of f) is not a unique function of the total plastic work per unit volume.

ACKNOWLEDGMENT.

We are grateful to the Director of the British Iron and Steel Research Association for permission to publish this work.

Fig. 2.



Stress–strain curves for copper at various inclinations to the rolling direction (experimental points omitted for clarity).

REFERENCES.

BALDWIN, W. M., Jr., HOWALD, T. S., and ROSS, A. W., 1946, *Trans. Am. Inst. Min. Met. Eng.*, **166**, 86.

BURGHOFF, H. L., and BOHLEN, E. C., 1942, *Trans. Am. Inst. Min. Met. Eng.*, **147**, 144.

CHEVIGNY, R., 1949, *Revue de l'Aluminium*, March, p. 79; 1950, *Sheet Metal Industries*, **27**, 133.

DORN, J. E., 1949, *Journ. App. Phys.*, **20**, 15.

HAZLETT, T. H., ROBINSON, A. T., and DORN, J. E., 1949, University of California, Engineering Research Project N6-ori-111, Task No. 4, May.

HILL, R., 1948, *Proc. Roy. Soc.* A, **193**, 281; 1949, British Iron and Steel Research Association, Report MW/E/48/49, Sept.; 1950, *The Mathematical Theory of Plasticity*, Chap. XII, (Clarendon Press: Oxford).

JACKSON, L. R., SMITH, K. F., and LANKFORD, W. T., 1948, *Metals Technology, Tech. Pub.* 2440, August; see also discussion of this paper in 1949, *Journal of Metals*, **1**, 324.

KLINGLER, L. J., and SACHS, G., 1948, *Journ. Aero. Sci.*, **15**, 599.

LANKFORD, W. T., SNYDER, S. C., and BAUSCHER, J. A., to be published in *Trans. Am. Soc. Metals.*

LXI. *The Fluctuation and Fading of Radio Echoes from Meteor Trails.*

By J. S. GREENHOW,
University of Manchester *,

[Received April 21, 1950.]

[Plates XXII. & XXIII.]

ABSTRACT.

This paper is concerned with the amplitude fluctuations observed in the radio waves reflected from meteor trails. Subsequent to the initial diffraction phenomena the radio echoes exhibit fluctuations of period between 0·01 and 0·1 seconds, when observed on radio wave frequencies of 36 to 72 Mc./s. These fluctuations have been investigated simultaneously on two radio wave frequencies and it is concluded that the phenomenon is due to wind gradients in the high atmosphere which cause the trail of meteor ionization to break up. In this way several reflecting regions are formed which give an interference pattern moving over the observing station.

§ 1. INTRODUCTION.

OBSERVATIONS of the radio echoes from meteor trails indicate that the amplitude fluctuations during the life of the echo are complex in character. For example, during the Giacobinid shower of October 1946, Lovell, Banwell and Clegg (1947) investigated the life histories of several echoes and found that many showed rapid fluctuations in amplitude superimposed on the general decay of the echo. Similar effects have been found by McKinley and Millman (1949). These observations were made using either cine-photography or photography of intensity modulated displays, and it was evident that further progress in the study of this phenomenon could only be made by investigating the amplitude of the echo over much shorter time intervals. This paper describes the information which has now been obtained on the fluctuations using a technique which enables the amplitude of successive radio pulses to be measured with a time separation of 6·6 milliseconds.

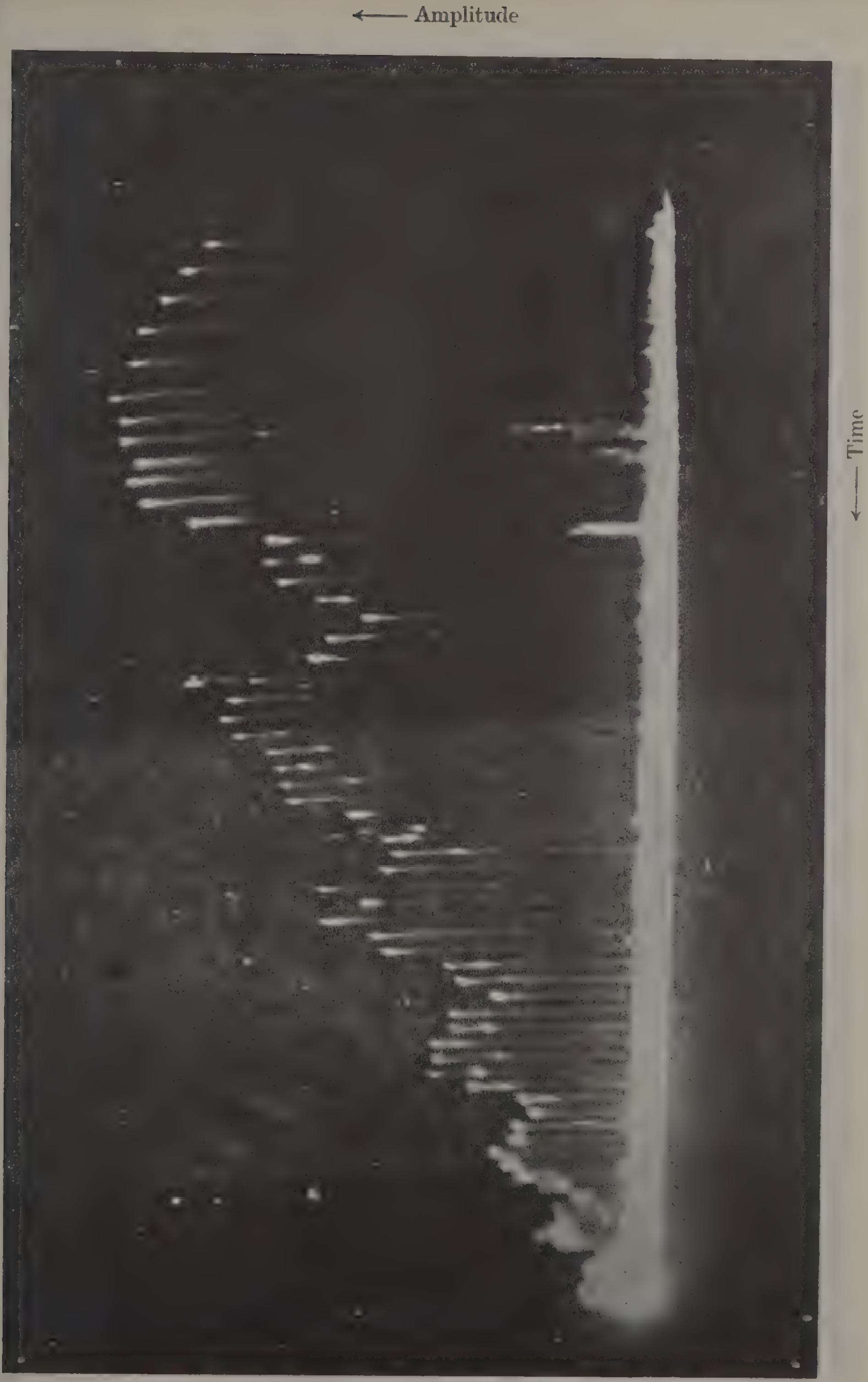
During the complete life history of a characteristic radio echo from a meteor trail it is possible to distinguish four types of fluctuation.

(*a*) *Diffraction Phenomena.*

Approximately 10 per cent of meteor echoes show a rapid initial rise in amplitude, followed by oscillations of increasing frequency. A typical echo of this type is shown in fig. 1 (Pl. XXII.). The oscillations, which are usually

---

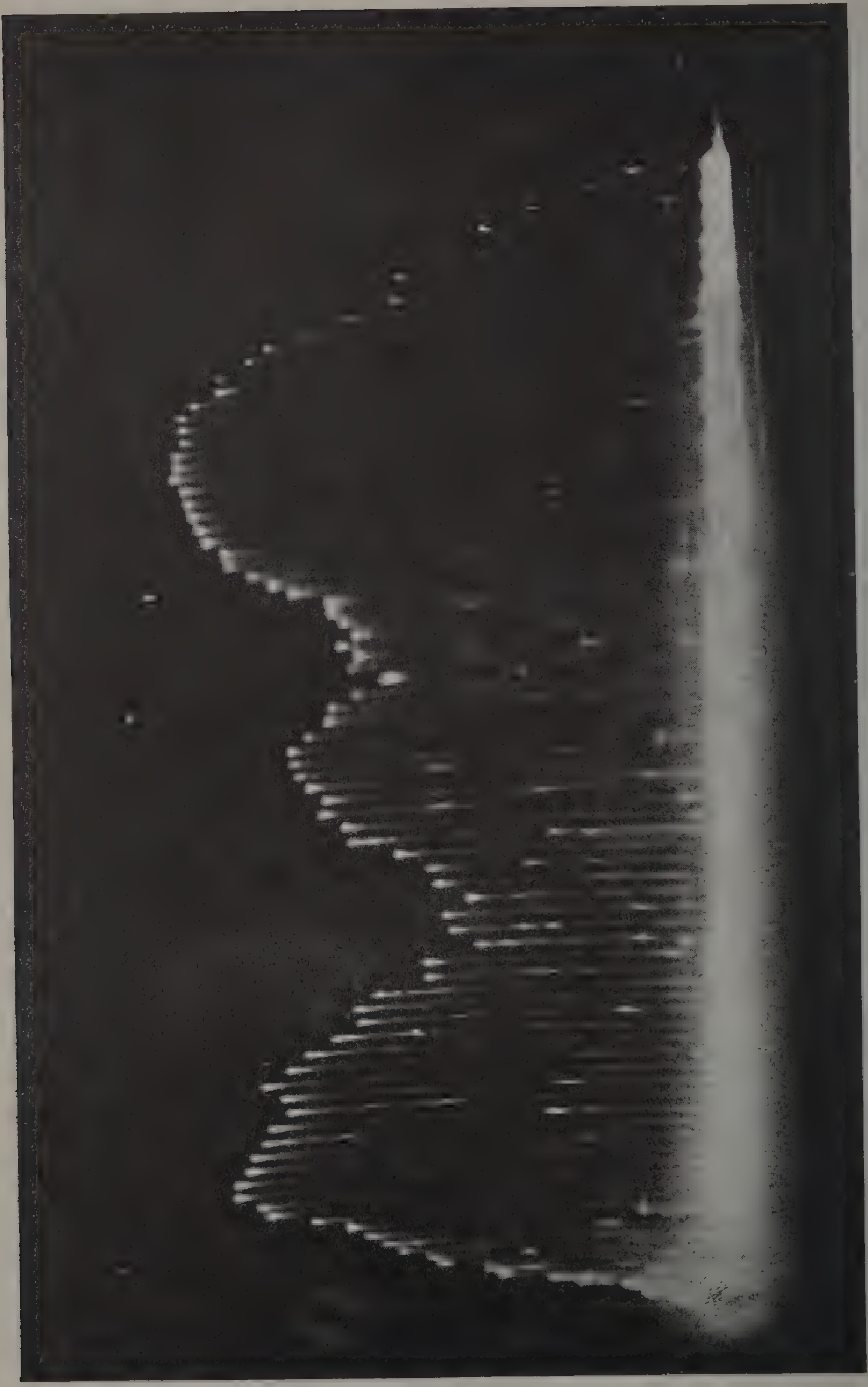
* Communicated by Dr. A. C. B. Lovell.



Typical diffraction pattern. Velocity of meteor: 340 km/sec.

To face page 682

Fig. 2.



Example of short period flutter. Pulse repetition frequency: 750 c/s. Mean Period of flutter: 0·04 seconds.

complete with 0·5 seconds of the first recorded pulse correspond to the passage of the head of the meteor through successive Fresnel diffraction zones, and they can be used to measure the meteor's velocity (Ellyett and Davies 1948, Davies and Ellyett 1949).

(*b*) *Short Period Flutter.*

Although the majority of echoes have durations of less than 0·2 sec., some 25 per cent persist for much longer periods, and show violent fluctuations of the type illustrated in fig. 2 (Pl. XXIII.). These oscillations, which have periods of the order 0·01 to 0·1 seconds, are occasionally observed directly following the Fresnel fluctuations, but more usually occur with echoes where the diffraction phenomena are not observed.

(*c*) *Long Period Fluctuations.*

In a small percentage of very long duration echoes, a much slower amplitude variation, of the order of 1·0 seconds period, has been detected.

(*d*) *Noise Fluctuations.*

In addition to the foregoing types of amplitude oscillation all echoes invariably show small fluctuations between adjacent pulses, which are due to random noise in the receiver (Ellyett 1949).

The phenomena described in (*a*) and (*d*) are now well understood, and will not be discussed further in this paper. The short period flutter (*b*) and long period fluctuations (*c*), however, present a number of interesting physical problems, and have been investigated in detail.

§ 2. TECHNIQUE.

Most of the results have been obtained by a dual wave technique, giving continuous pulse by pulse amplitude records on 72 and 36 Mc./s. simultaneously, for the whole duration of an echo. Additional information has been obtained (on 72 Mc./s. only), from earlier techniques, using automatic photography of successive pulse amplitudes from sections of the echo.

(*a*) *Dual wave technique.*

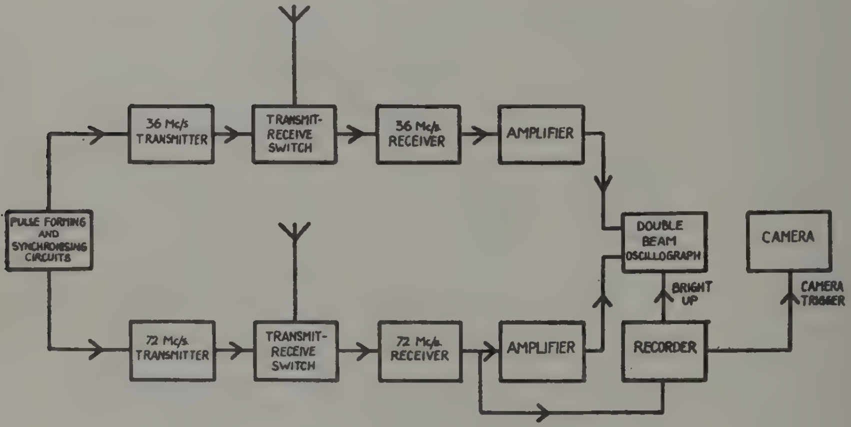
This technique was developed by Ellyett * and a block diagram of the equipment is shown in fig. 3. Two transmitting and receiving systems were used, operating at frequencies of 72 and 36 Mc./s. Upon receipt of the first pulse exceeding twice the signal to noise ratio reflected from a meteor trail, a recorder unit (Davies and Ellyett 1949), attached to the 72 Mc./s. receiver output, actuated the dual wave unit. Both receiver outputs were fed continuously via separate amplifying channels in this unit, to a double beam cathode ray tube. A camera with continuously moving film was switched on simultaneously, so that successive pulses were separated on the film, giving a pulse by pulse picture of the echo amplitude throughout its life. The camera motor required approximately 0·05 sec. to reach a speed sufficient to separate the individual pulses, so this period was lost at the beginning of each echo. Once the recorder had actuated the system a hold-on relay came into operation, maintaining the unit in action, until switched off manually on conclusion of the echo.

---

* Dr. C. D. Ellyett returned to New Zealand before the final completion of the dual wave technique.

The two transmitters were locked together to the same pulse repetition frequency, and their peak powers were adjusted so that echoes appearing on both wavelengths simultaneously had approximately the same signal to noise ratios. The aerials were horizontal folded dipoles with reflectors, at heights of half a wavelength above the ground, and looking in a northerly direction. Common transmit-receive systems enabled the same aerials to be used for both transmitting and receiving. The pulse repetition frequency used in these experiments was 150 per sec.

Fig. 3.


36 Mc/s TRANSMITTER
TRANSMIT-RECEIVE SWITCH
36 Mc/s RECEIVER
AMPLIFIER
PULSE FORMING AND SYNCHROISING CIRCUITS
DOUBLE BEAM OSCILLOGRAPH
CAMERA
BRIGHT UP
CAMERA TRIGGER
72 Mc/s TRANSMITTER
TRANSMIT-RECEIVE SWITCH
72 Mc/s RECEIVER
AMPLIFIER
RECORDER

Block diagram of the dual wave apparatus.

(*b*) *Single wave technique.*

An automatic recorder unit designed for observation of the Fresnel zone pattern has been described previously (Davies and Ellyett 1949). This unit photographs some 100 pulses on one photographic frame, and takes 1 sec., to reset. For long duration echoes detailed amplitude information can therefore be obtained from short sections of the echo separated by intervals of 1 second.

§ 3. The Short Period Flutter.

Photographic records of short period flutter have been obtained during most of the major showers of 1948 and 1949, although the dual wave technique was used only during the 1949 Geminid shower. Examination of the results shows that the individual fluctuations which comprise the flutter vary continuously in amplitude and frequency during the lifetime of an echo. The results have been analysed to find, (*a*) the relationship between the echo duration and the mean time period of the flutter ;

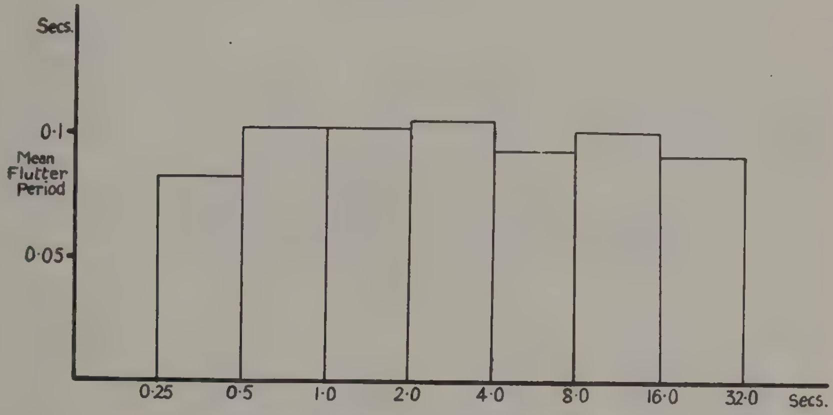
(*b*) the relationship between the mean flutter periods observed on frequencies of 36 Mc./s. and 72 Mc./s. ; and

(*c*) the relationship between the flutter period and the type of meteor.

(*a*) *Relation between Echo duration and Mean Period of Flutter.*

A histogram relating the mean period of the flutter to the echo duration, for echoes observed on a frequency of 36 Mc./s., is shown in fig. 4. It is clear that the mean period is practically the same for each duration group. The processes causing the phenomenon of high speed flutter must therefore be independent of the duration of the echo.

Fig. 4.



Echo duration. (36 Mc./s.).
Mean flutter period as a function of echo duration.

(*b*) *Relationships between the Mean Periods of Fluctuation on frequencies of* 38 *Mc./s. and* 72 *Mc./s.*

The relationship between the flutter periods on 36 Mc./s. and 72 Mc./s. was investigated with the aid of the dual wave apparatus.

Table I. gives the mean flutter periods for the two frequencies, and the ratios between these periods.

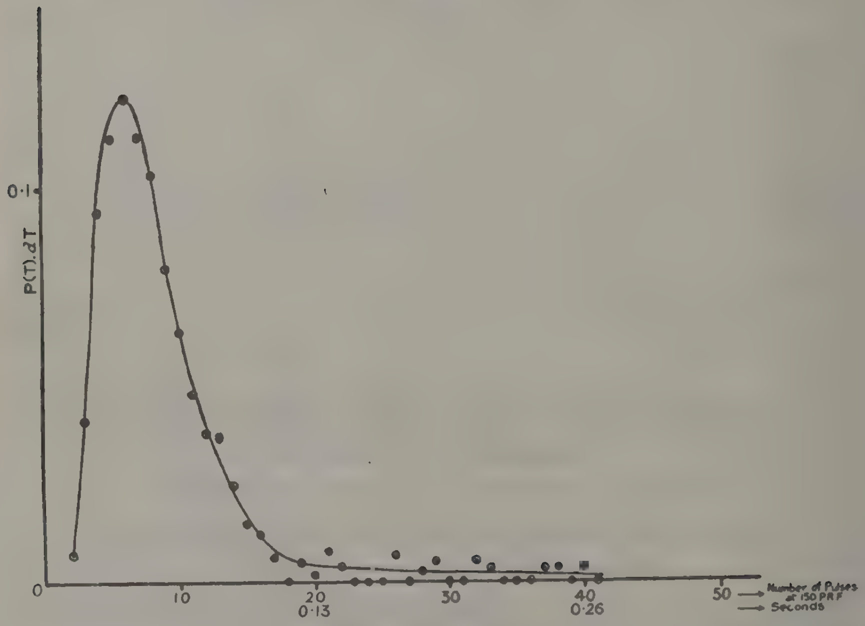
TABLE I.

	Mean flutter period (secs.)	Ratio of the mean flutter period at 36 Mc./s. to the mean flutter period at 72 Mc./s.
72 Mc./s. echoes	0·045	1·0
36 Mc./s. echoes while accompanied by a 72 Mc./s. echo	0·082	1·84
All 36 Mc./s. echoes	0·091	2·02

It is evident that the period of the oscillations on a frequency of 36 Mc./s. is approximately twice that of those observed on 72 Mc./s. In general the echo observed on a wavelength of 4 m. is of much shorter duration than the echo observed on 8 m. from the same trail (Lovell 1948) but there is no significant difference between the mean periods on 36 Mc./s. before and after the end of the corresponding 72 Mc./s. echo.

Since the flutter period is independent of the duration of the echo, the probability P(T) of a fluctuation of period T occurring is proportional to T . N(T), where N(T) is the total number of fluctuations of this period observed. The relationship between T . N(T) and T for the echoes photographed on the 72 Mc./s. apparatus is shown in fig. 5. This is compared with the corresponding curve for 36 Mc./s. echoes in fig. 6.

Fig. 5.



Flutter period (T).

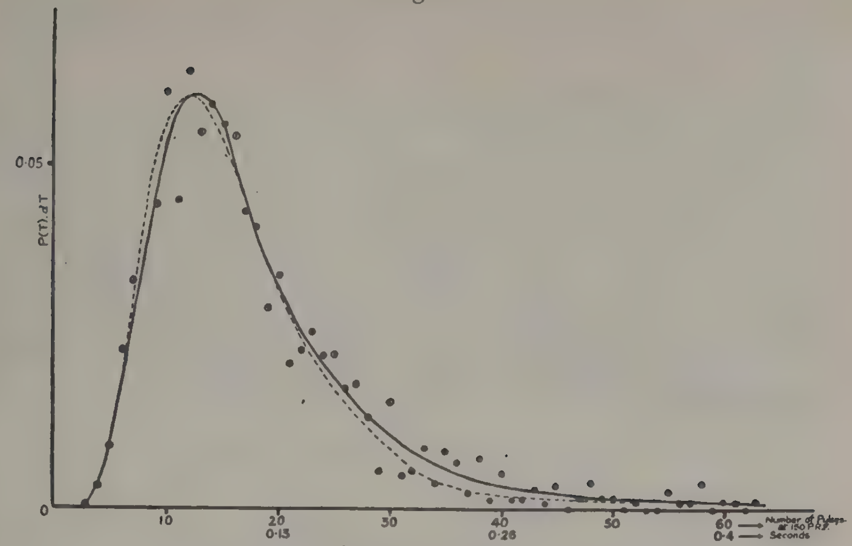
P(T)*d*T: Probability of a fluctuation of period between T and T+*d*T occurring at 72 Mc./s. (*d*T=0·066 secs).

If the mean flutter periods T_1 and T_2 observed on the 36 Mc./s. and 72 Mc./s. equipments are related to the wavelengths by a power law, the exponent n will be given by

$$n = \frac{\log_{10} T_1/T_2}{\cdot 301}.$$

This exponent was found for 36 echoes observed simultaneously on both wavelengths. Measurements were restricted to times when both echoes were present, and the distribution of values obtained for n is shown in

Fig. 6.

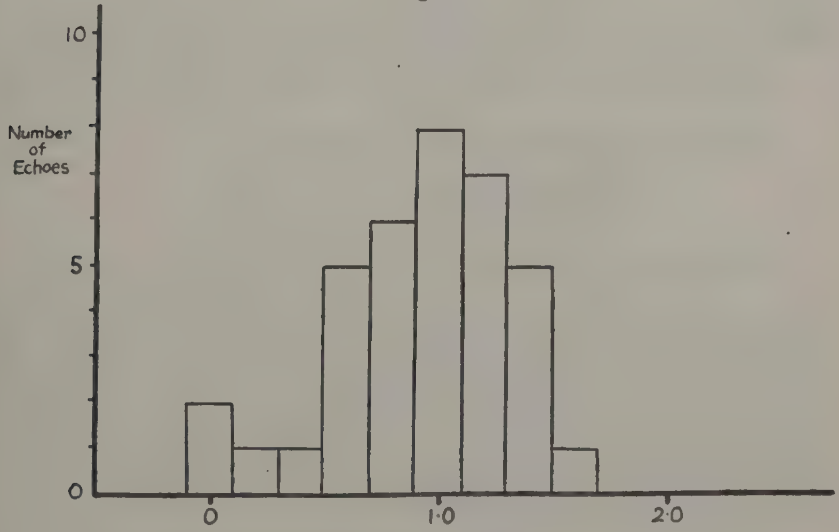


Flutter period (T).

(Continuous curve). P(T)*d*T : probability of a fluctuation of period between T and T+*d*T occurring at 36 Mc./s. (*d*T=0·066 secs.)

(Dotted curve). Corresponding curve for 72 Mc./s. scaled up by a factor of 2.

Fig. 7.



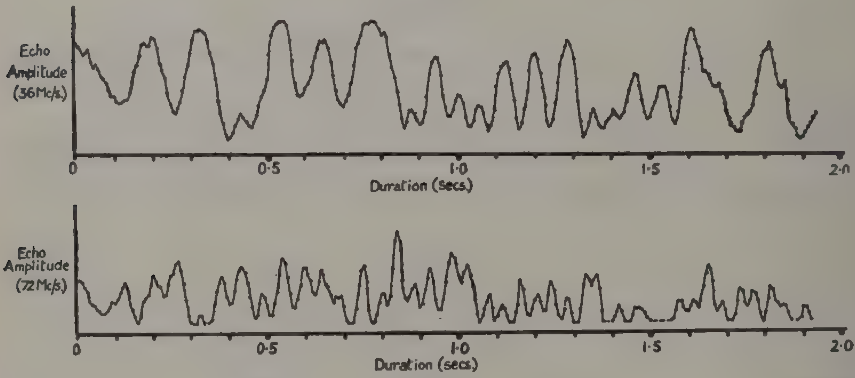
Exponent (n) calculated from $\frac{T_1}{T_2} \doteq \left(\frac{\lambda_1}{\lambda_2}\right)^n$

T_1 and T_2 are mean flutter periods at wavelengths of λ_1 and λ_2.
Distribution curve for echoes present simultaneously on wavelengths of 8·4m. (λ_1) and 4·2m. (λ_2).

fig. 7. The mean value was found to be 0·99. It therefore seems reasonable to conclude that the periods are proportional to the wavelength of the radiation employed.

Thirty-six echoes exceeding 0·5 sec., in duration at 72 Mc./s., whose amplitudes remain at a reasonable value for a sufficiently long period, have been analysed pulse by pulse. In general, the 72 Mc./s. fluctuations are of shorter periods than those at 36 Mc./s., and there are sections in most of the echoes where the ratio of the periods is exactly two to one. Fig. 8 shows an echo in which a high degree of correlation persists for an appreciable part of the echo life. In this example amplitude maxima at the higher frequency correspond approximately to the maxima and minima at 36 Mc./s. The echo lasted for 1·9 sec. at 72 Mc./s. and 5·8 secs. at 36 Mc./s., only the first 2·0 secs. of the 36 Mc./s. echoes are shown. The individual pulse amplitudes have been smoothed by a sliding group of two.

Fig. 8.



Time: 05.21 hrs. Dec. 13th 1949.

Pulse repetition frequency 150 c/s.
Duration of 72 Mc./s. echo 1·9 secs.
Duration of 36 Mc./s. echo 5·8 secs.

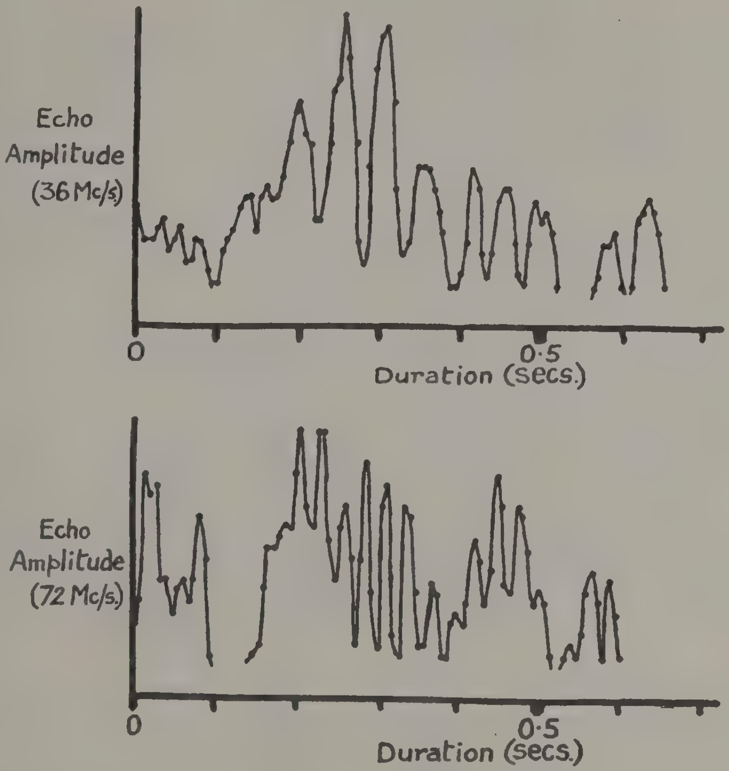
Individual pulse amplitudes smoothed by a sliding group of 2.

Fig. 9 is an echo of shorter duration showing higher speed fluctuations. The pulse amplitudes are unsmoothed. Fig. 10 is an example of an echo in which the fluctuations are less clearly defined. There is a two to one ratio in flutter periods for certain sections of the echo, but in other parts the fluctuations are confused and often of small amplitude. In general the fluctuations at 72 Mc./s. are faster than those at 36 Mc./s. The first 3·9 sec., of the echo are shown on both frequencies, although the 36 Mc./s. echo persists for a further 14 sec.

(c) *Relation between flutter period and types of meteor.*

In order to discover whether the flutter period is related to the type of meteor by which the trail was originally produced, or if it depends only upon the subsequent behaviour of the completed column of ionization, the mean flutter periods for three major showers, and for sporadic meteors have been measured on 72 Mc./s. only, using the single wave technique. The results are given in Table II.

Fig. 9.



Time 22.24 hrs. Dec. 13th 1949.

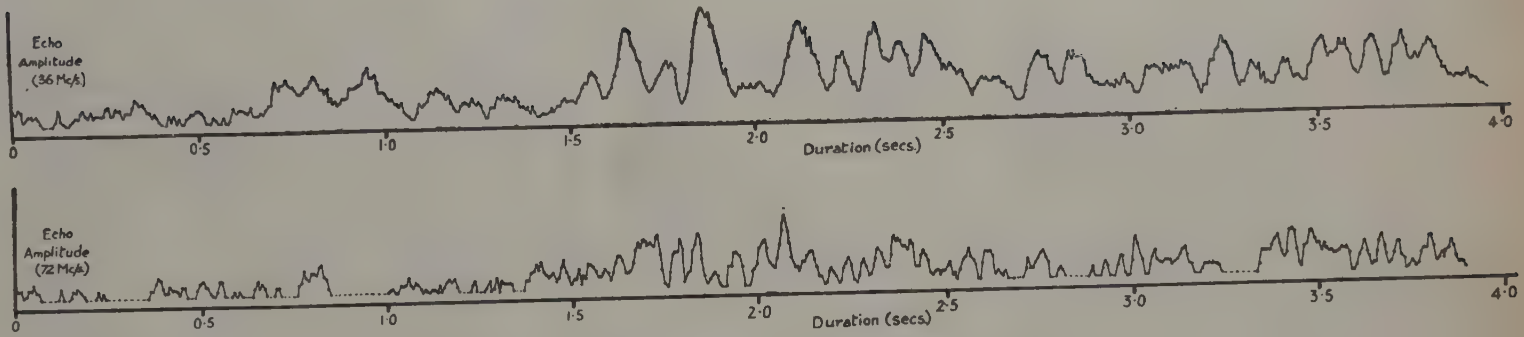
Pulse repetition frequency	150 c/s.
Duration of 72 Mc./s. echo	0·6 secs.
Duration of 36 Mc./s. echo	2·0 secs.

TABLE II.

Mean Period of Flutter (72 Mc./s.).

Shower	No. of observations of period	Mean period of flutter (sec.)
Perseids (Aug. 1948)	80	0·048
Geminids (Dec. 1948)	362	0·032
Quadrantids (Jan. 1949)	74	0·035
Perseids (Aug. 1949)	353	0·039
Sporadic meteors	41	0·033

Fig. 10.



Time : 0046 hrs Dec. 13th 1949.
Pulse repetition frequency 150 c/s.
Duration of 72 Mc./s. echo 3·9 secs.
Duration of 36 Mc./s. echo 17·1 secs.

It is evident that the mean period is independent of the type of meteor which produces the initial column of ionization.

§ 4. Long Period Oscillations.

Some experimental evidence for the existence of a longer period amplitude fluctuation has been found on 36 Mc./s. only. Out of a total of 30 echoes obtained by Ellyett * during the 1949 Perseid meteor shower, two, with durations of 81·0 sec. and 50·0 sec., show long period fluctuations emerging after about 30·0 sec., the period of the fluctuations gradually increases throughout the remainder of the echo life, and the mean period is of the order of 1·0 sec.

Similar behaviour has not yet been found using the dual wave equipment, owing to the shorter durations of the echoes on 72 Mc./s., but evidence of waves in one 7·0 sec. echo on 72 Mc./s. has been given by Lovell, Banwell and Clegg (1947), using cine-photography. Further information on the behaviour of the 72 Mc./s. echoes was obtained from a total of 90 cine records of the summer daytime and Perseid meteors of 1947, which show 10 cases of possible wave occurrence. A detailed analysis of the periods of the oscillations cannot be made from these records, owing to the possibility of a beating effect with the camera shutter mechanism.

As there are only two positive cases of the occurrence of long period amplitude fluctuations at 36 Mc./s., and none at 72 Mc./s., it is not yet possible to decide whether this is a separate phenomenon, or merely the tail of the distribution curve shown in fig. 4.

§ 5. Theoretical Explanations of the High Speed Amplitude Fluctuations.

Experimental evidence from several sources (for summary see Ellyett 1950) shows that there are both steady winds, with velocities of the order of 50–100 m./sec., and winds of a turbulent character, present in the region 70–120 km. above the earth.

Several suggestions have been made that the phenomenon of high speed flutter might be explained by the presence of such wind gradients in the upper atmosphere. (Herlofson 1948, Ellyett 1949, McKinley and Millman 1949). The results obtained in this paper will now be considered in terms of such wind motions. When a meteor trail is first formed, the ionization along the track is approximate linear, and it is known that the reflection of radio waves at 72 Mc./s. from the trail is specular (Lovell, Banwell and Clegg 1947, Hey and Stewart 1949). If the trail is subsequently deformed by winds, then other sections could present right angle reflecting points to the observer. An interference pattern due to the interaction of waves

---

* Unpublished.

reflected from various sections of the trail will then be formed on the surface of the earth. If the reflecting centres move relatively to each other, this interference pattern will move over the ground, producing a series of maxima and minima in the echo amplitude at any point.

If two sections of the trail predominate, and reflect approximately equal amounts of energy in the direction of the receiver, large amplitude fluctuations will be produced. The frequency of the fluctuations will be proportional to the frequency of the radio waves. The maximum possible flutter frequency will occur when the reflecting points are moving in opposite directions along the line of sight of the observer, with the maximum wind velocities present in the upper atmosphere. Any other flutter frequency less than this maximum is possible.

If V is the component of the velocity of the sources relative to each other in the direction of the observer, λ the wavelength, and T the period of the flutter, then

$$\mathrm{V} \sim \frac{\lambda}{4\mathrm{T}}.$$

V is independent of the wavelength, so that the flutter period T should be proportional to λ.

The distribution curve for all flutter periods at 72 Mc./s., scaled up by a factor of two is shown together with the corresponding distribution for 36 Mc./s. in fig. 2. (Pl. XXIII.). There is a reasonable correlation between the two curves. The most probable value of V is approximately 25 metres/sec., while the highest speed fluctuations correspond to a value of V of the order of 100 metres/sec. The slowest fluctuations observed have periods of approximately 0·5 sec. at 36 Mc./s. and 0·25 sec. at 72 Mc./s.

If more than two reflecting centres, of differing amplitudes, and moving with different velocities, are present, the fluctuations will become more complex, and correlations on widely different frequencies could not be expected. The mean flutter period, however, would still be proportional to the wavelength.

Additional information supporting the wind theory has been obtained by the visual observation of an auxiliary cathode ray tube. In several cases an echo was observed at 36 Mc./s. a few seconds before the appearance of the 72 Mc./s. echo. This suggests strongly that the track is undergoing distortion, and that a favourable electron distribution in the trail has temporarily caused the 72 Mc./s. echo to appear.

ACKNOWLEDGMENTS.

The author wishes to thank the Director and Staff of the Manchester University Jodrell Bank Experimental Station, where these experiments were carried out; especially Dr. C. D. Ellyett, who analysed most of the single wave records and the cine records referred to in § 4, and Mr. R. L. Closs, who assisted materially in the taking of the dual wave observations. Valuable advice was given by Dr. A. C. B. Lovell, Director of the Station, and by Dr. J. A. Clegg.

Much of the work described in this paper has been made possible by the grant from the Department of Scientific and Industrial Research for the development of the experimental work at Jodrell Bank. The author is also indebted to the D.S.I.R. for the financial aid given during the course of this work.

REFERENCES.

DAVIES, J. G., and ELLYETT, C. D., 1949, *Phil. Mag.*, **40**, 614.

ELLYETT, C. D., 1949, *Ph.D. Thesis*, University of Manchester; 1950, *Phil. Mag.*, **41**, 694.

ELLYETT, C. D., and DAVIES, J. G., 1948, *Nature, Lond.*, **161**, 596.

HERLOFSON, N., 1948, *Rep. Prog. Phys.*, **11**, 444 (London: Physical Society).

HEY, J. S., and STEWART, G. S., 1947, *Proc. Phys. Soc.*, **59**, 858.

LOVELL, A. C. B., 1948, *Rep. Prog. Phys.*, **11**, 415 (London: Physical Society).

LOVELL, A. C. B., BANWELL, C. J., and CLEGG, J. A., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 164.

McKINLEY, D. W. R., and MILLMAN, P. M., 1949, *Proc. Inst. Radio Engrs.*, **37**, 364.

LXII. *The Influence of High Altitude Winds on Meteor Trail Ionization.*

By C. D. ELLYETT, Ph.D. *,
University of Manchester †.

[Received April 21, 1950.]

ABSTRACT.

The paper contains a discussion of the possible effect of high altitude winds on the ionized columns created by meteors. It is shown that the various types of range drifts observed in the long duration radio echoes from meteor trails can be explained in terms of the influence of winds on the ionized column. A comparison of the characteristics of the fading of the radio waves returned from the ionosphere, with the characteristics of the fluctuations of the radio echoes from meteor trails, leads to the conclusion that some of the ionospheric phenomena may be due to meteor ionization in the E region.

§ 1. INTRODUCTION.

THERE is now much evidence from the study of the radio wave reflections from meteor trails to show that the trail undergoes complex disturbances after its formation, and several authors have suggested that winds at altitudes of 80–120 km. are an important disturbing factor (Herlofson 1947, Hey and Stewart 1947, McKinley and Millman 1949). The purpose of the present paper is to examine the available evidence and to estimate the extent to which winds can account for the observed phenomena. The principal phenomena involved are those of range drifts and amplitude fluctuations found in the long duration echoes from meteor trails. The paper includes a discussion of the correlation of these effects with recent work on the fading of radio waves from the ionosphere (Ratcliffe 1948).

§ 2. WIND PROCESSES AT 70–100 km. HEIGHT.

Evidence from the movement of polar auroræ, of noctilucent clouds, and of meteor dust trails seen by reflected sunlight, gives conclusive proof of wind motion at heights of 70–100 km. (Störmer 1939). The mean drift velocity is of the order of 50–100 m./sec. Similar values for the wind velocities have been obtained recently by Mitra (1949) from direct ionospheric radio measurements.

In addition to steady winds, turbulent motion also occurs. Olivier (1947), analysing the movements of 1492 visual and telescopic meteor trails, finds proof of winds with different velocities and directions along a given trail, and both vertical and horizontal components within quite small volumes.

---

* Communicated by Dr. A. C. B. Lovell.

† Now at Canterbury University College, New Zealand.

Based on these observations, a trail of meteor ionization, initially linear, may be subject to one or more of the following types of motion :—

(i) Drift or tilt of the trail as a complete unit ;
(ii) Gradual curvature of large sections of the trail ;
(iii) Relative movement of short sections of the trail, due to turbulence.

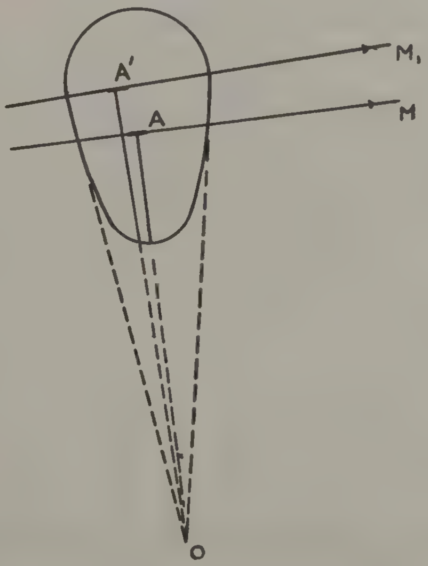
These effects will be considered in turn.

§ 3. INTERPRETATION OF OBSERVATIONS IN TERMS OF WIND MOTION.

(i) *Range Drift—Drift or tilt of the trail as a complete unit.*

It is shown by the work of Lovell, Banwell and Clegg (1947), and by Hey and Stewart (1947) that the reflection from a meteor trail is markedly aspect sensitive at 72 Mc./s. In fig. 1, if M is the line of motion of the meteor, most of the echo energy received at O will come from the region close to the foot A of the perpendicular from O to the trail. The elliptical curve represents the coverage of the aerial over a horizontal plane at a height of 100 km. The meteor trails which give echoes are contained completely within this area*.

Fig. 1.



If the trail drifts as a unit nearly parallel to itself to a position M_1, under the influence of winds, the reflecting region will move to A′. The velocity of such movement should be of the order of 50 m./sec., and could be either towards or away from the receiver. The change of range in drifts of this order is very slight, being only 1·25 km. in 25 sec. Drifts of this magnitude have been found by McKinley and Millman (1949) who report that the majority of long-enduring discrete range echoes exhibit a drift in slant range at the rate of approximately 1·0 km. in 25 sec. Their result is in good numerical agreement with the hypothesis of a general wind drift. The further observation by McKinley and Millman that the drift is predominantly away from the observer still awaits explanation.

---

* The polar diagram has a volume distribution for meteor collection (Clegg 1948), but for present illustrative purposes the section of this volume is sufficient.

(ii) *Range Drift—Gradual curvature of large sections of the trail.*

In addition to this slow drift, much greater range movements have been seen in a proportion of echoes. The available experimental evidence on these gross range drifts observed on 72–73 Mc./s. is given in Table I.

Table I.

Observations on Range Drift.

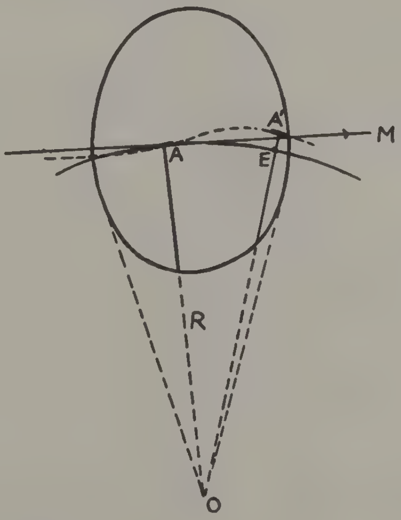
Equipment				Range drifts			
Author	Freq. (Mc./s.)	Aerial array	Beam width	Total No. of echoes examined	No. of travelling echoes	Order and sign of movement	Velocity of movement
Prentice, Lovell, and Banwell 1947	72	*Vertical* Yagi	Main beam 21°. Appreciable side lobes	1836	55·0 (3%)	5 to 10 km. Mostly decreasing range	Widely spread from a few km./sec. to 60 km./sec.
Hey, and Stewart 1947	73	*Vertical* Yagi	15°	—	2·0%	Mostly decreasing range	0 to 25 km./sec.
Ellyett *	72	Array of 5 Yagis at 15° elevation	Az. 6·5° El. 14°	>1000	(0·1%)	+ 7·0 km.	+ 0·39 km./sec.

* Unpublished. From an analysis of more than 1000 cine records obtained during the major showers of 1946–48 using the technique described by Lovell, Banwell and Clegg (1947).

The first point of interest emerging from the table is the agreement between the three independent sets of observations with respect to the magnitude and the velocity of apparent movements, which are of a very high order; speeds of 25 km./sec. being by no means exceptional. The suggestion was made by Hey and Stewart (1947) that since these drift velocities were far in excess of any possible direct air movement, the distortion of the trail must have produced a change in the point of reflection, leading to the very high apparent rate of movement. Type (ii) of the classes of air motion suggested is identical with this mechanism. Referring to fig. 2 the initial reflecting region of the trail at A is shifted to A′ by progressive distortion. The actual displacement of the trail is exaggerated in the figure, the region A′ being visualized as lying very close to its original position along M. The one definite range drift obtained from an analysis of the cine records (Table I.) is the only echo available with sufficient information to allow an estimate to be made of the wind gradient. As the range R=167 km. and the drift EA′=+7 km. in 18·0 sec., the angular displacement EOA is 16°. This yields a wind velocity gradient of 9 m./sec./km. This value, being less than one-fifth of the mean wind velocity, appears not unreasonable.

On the mechanism of fig. 2, since reflection effectively comes from some new point along the line M, the range always increases, and must do so in all cases where the normal to a linear trail lies within the beam area for initial reflection.

Fig. 2.

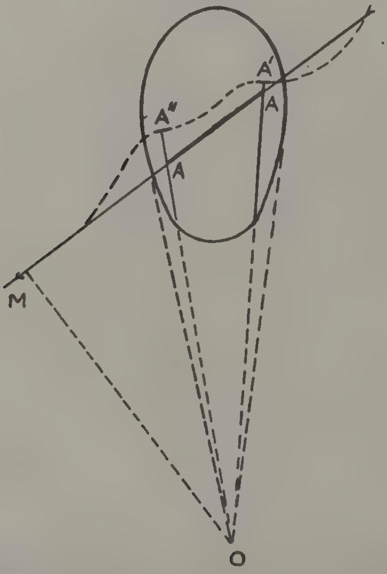


Although a percentage of outward drifts do occur, the major group appears to be directed towards the observer. To account for this group a slightly different mechanism must be postulated. Consider a meteor track, given by the heavy section AA in fig. 3, passing obliquely through the aerial beam, so that the projected perpendicular reflecting point lies outside the beam angle. In such a case no echo is received on 72 Mc./s. unless the track undergoes distortion and presents a suitable directed portion A′ within the beam. As the bending increases A′ may decrease in range towards A″.

An interesting feature of the results given in Table I. is the marked difference between the numbers of echoes showing range drift with vertically and horizontally directed beams. This can also be explained on the concept of trail distortion. With a vertically directed beam an echo can only be obtained by specular reflection if the meteor trail is almost horizontal, or if the meteor emanates from a radiant close to the horizon. Under these circumstances the rate will be low (Prentice 1948). Hence a larger proportion of the echoes observed with a vertical beam are due to trail distortion. A simple arithmetical estimate, taking into account the relative collecting areas (Clegg 1948) shows that on this hypothesis, the percentage of range drifts observed with a narrow vertical beam should exceed that observed with a horizontal beam by a factor of the order of 20. This agrees well with the ratio obtained experimentally.

As the radio wavelength is lowered the echo duration increases, and hence the occurrence of reflection at angles other than the normal to the initial trail should increase. This has been substantiated by the work of McKinley and Millman at 32·7 Mc./s., who find that trails occurring visually within ±10° of a direction perpendicular to the line of sight from the station to the trail produce radio echoes after a delay of about 1·0 sec., whereas for angles of departure lying between 50° and 90° the delay increases to 10·0 sec.

Fig. 3.



(iii) *Turbulent Motion—Relative movement of short sections of the trail.*

The third process considered in § 2 concerned the relative movements of short sections of the trail. The explanation of the high speed amplitude fluctuations in terms of this type of trail movement has been discussed by Greenhow (1950). In the presence of such turbulence the aspect sensitivity of the trail for radio reflections should slowly vanish, and appreciable range spreading of long duration echoes should gradually occur, being more marked on the lower radio frequencies. Evidence for this effect has been obtained on 72 Mc./s. (Prentice, Lovell and Banwell 1947), while McKinley and Millman 1949, have found marked evidence of the effect on 32·7 Mc./s.

§ 4. CORRELATION OF METEOR TRAIL FLUCTUATIONS WITH DIRECT OBSERVATIONS OF IONOSPHERIC FADING.

Studies of the fading of radio waves returned from the ionosphere suggest the presence of moving ionospheric irregularities (Ratcliffe 1948). These irregularities are considered to be caused by wind motion (Pawsey 1935, Findlay 1949, Mitra 1949). These measurements have been made at frequencies ranging from 16 Kc./s. to 6 Mc./s., and at such frequencies meteor ionization from individual trails will remain an effective echo source for periods of many minutes.

TABLE II.

Period of fading due to Relative Movement of Ionized Air at Heights above 80 km.

Method	Author	Period of fading	Equivalent period at 72 Mc./s.	
			Short	Long
Ionospheric echoes				
Amplitude fading curves	Ratcliffe (1949)	5/min. at 3 Mc./s. (approximate mean from spread of results)		0·7 sec.
Phase fading	Findlay (1949)	1/sec. at 2·4 Mc./s.	0·03 sec.	
Phase and amplitude fading	Bracewell (1948)	0·2/min. at 16 Kc./s. 0·67/hr. at 16 Kc./s.	0·07 sec.	1·2 sec.
Meteor trail echoes				
Amplitude flutter	Greenhow (1950)	0·04 sec. at 72 Mc./s. 0·09 sec. at 36 Mc./s.	0·04 sec. 0·04 sec.	
Longer period fading	Greenhow (1950)	1·0 sec. at 36 Mc./s.		0·5 sec.
Amplitude flutter	McKinley and Millman (1949)	5–10/sec. at 32·7 Mc./s.	0·09–0·05 sec.	

It has already been suggested by several authors that sporadic E effects are in part due to meteor ionization (summary by Lovell 1948). Evidence will now be produced which indicates that some of the irregularities causing the fading of radio waves from the ionosphere may also be due directly to wind distortion of meteor trails.

Since no measured wind velocities are yet available from meteor studies, the mean period of fading must be used as the point of comparison. In order to allow comparison of the results of various workers, it will be assumed that the speed of fading is proportional to the radio frequency employed. This assumption has already been made by Ratcliffe (1948), who has found some evidence of its validity. Further experimental evidence supporting this assumption has now been obtained by observation of the fluctuations occurring in the amplitude of radio waves reflected from meteor trails on 36 and 72 Mc./s. sumultaneously (Greenhow 1950).

Two distinct periods of ionospheric fading have been recognized on the very long wavelengths (Bracewell 1948) and although a large scaling up in frequency is required it is remarkable that the figures at 72 Mc./s. agree closely both with the short period flutter and with the longer period observed from meteor trails. A recent figure quoted by Ratcliffe at intermediate frequencies agrees with the longer period, and the figure given by Findlay (1949) is almost identical with the meteor flutter period.

There appears, then, to be a close correspondence in the flutter observed from meteor trails and the ionospheric fading obtained from unspecified " centres ". Both show similar characteristics and have the same mean period. It therefore seems not unlikely that some of the ionospheric fading is, in fact, directly due to meteor ionization at E region heights.

In the absence of further results, no great significance can as yet be attached to the correspondence of the longer periods.

ACKNOWLEDGMENTS.

The work described in this paper was carried out at the Jodrell Bank Experimental Station of the University of Manchester. Valuable advice was given by Dr. J. A. Clegg, and particularly by Dr. A. C. B. Lovell, Director of the station. The author is personally indebted to the Imperial Chemical Industries, Ltd., for the award of a Research Fellowship, and to the Council of Canterbury University College, Christchurch, New Zealand, for the granting of three years leave of absence.

REFERENCES.

BRACEWELL, R. N., 1948, *Nature, Lond.*, **162,** 9.

CLEGG, J. A., 1948, *Phil. Mag.*, **39,** 577.

FINDLAY, J. W., 1949, Private Communication.

GREENHOW, J. S., 1950, *Phil. Mag.*, **41,** 682.

HERLOFSON, N., 1947, *Nature, Lond.*, **160,** 74.

HEY, J. S., and STEWART, G. S., 1947, *Proc. Phys. Soc.*, **59,** 858.

LOVELL, A. C. B., BANWELL, C. J., and CLEGG, J. A., 1947, *Mon. Not. R. Astr. Soc.*, **107,** 164.

LOVELL, A. C. B., 1948, *Rep. Prog. Phys.*, **11,** 415 (London : Physical Society).

McKINLEY, D. W. R., and MILLMAN, P. M., 1949, *Proc. Inst. Radio Engrs.*, **37,** 364.

MITRA, S. N., 1949, *Proc. Inst. Elect. Engrs.*, Pt. III, **96,** 441.

OLIVIER, C. P., 1947, *Proc. Amer. Phil. Soc.*, **91,** 315.

PAWSEY, L. J., 1935, *Proc. Camb. Phil. Soc.*, **31,** 125.

PRENTICE, J. P. M., LOVELL, A. C. B., and BANWELL, C. J., 1947, *Mon. Not. R. Astr. Soc.*, **107,** 155.

PRENTICE, J. P. M., 1948, *Rep. Prog. Phys.*, **11,** 389 (London : Physical Society).

RATCLIFFE, J. A., 1948, *Nature, Lond.*, **162,** 9 ; 1949, Private communication.

STÖRMER, C., 1939, *Astrophysica Norwegica*, **3,** 117.

LXIII. *Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.*—Part V. *The Neutral Mesons.*

By A. G. CARLSON *, J. E. HOOPER and D. T. KING,
The H. H. Wills Physical Laboratory, University of Bristol †.

[Received May 24, 1950.]

[Plate XXIV.]

SUMMARY.

The spectrum of the γ-radiation in the atmosphere at 70,000 ft. has been determined by observations on the scattering of pairs of fast electrons recorded in photographic emulsions exposed in high-flying balloons. The detailed form of the spectrum is consistent with the assumption that the γ-rays originate by the decay of neutral mesons. It is found that the mass of the neutral mesons is 295 ± 20 m_e, and that they are created in nuclear explosions with an "energy spectrum" similar to that of the charged π-particles.

A method is described for determining the lifetime of the neutral mesons and their frequency of occurrence compared with charged π-particles. The lifetime, τ_{π^0}, is less than 5×10^{-14} sec. It may be possible to determine the lifetime, by observations of greater statistical weight, if it is longer than 2×10^{-14} sec. The ratio of the number of neutral mesons to charged mesons, produced in nuclear explosions of great energy, is equal to $0{\cdot}45 \pm 0{\cdot}10$.

1. INTRODUCTION.

RECENT experiments by Bjorklund, Crandall, Moyer and York (1950), strongly suggest that the high energy γ-rays observed when matter is bombarded by protons with an energy greater than about 200 MeV. are of secondary origin; that they arise by the decay into two quanta of neutral mesons, of mass about 300 m_e and lifetime less than 10^{-11} secs., which are the primary products of the nuclear collisions. Further, Bradt, Kaplon and Peters (1950), have made a detailed analysis of a disintegration produced in a photographic emulsion by an α-particle with an energy between 10^{12} and 10^{13} eV., which leads to the emission of fifty-seven shower particles. In the "core" of the shower particles, most of which are mesons (Piccioni 1950, Fowler 1950), these authors observe the production of several pairs of electrons, and conclude that the charged mesons are accompanied by γ-radiation. If this radiation is also created by the decay of neutral mesons—and if the latter particles are produced in the nuclear explosion with values of the kinetic energy similar to those of the charged mesons—it may be shown that the proper lifetime of the neutral mesons is less than 3×10^{-13} secs.

---

* On leave from the University of Upsala.

† Communicated by Professor C. F. Powell, F.R.S.

Finally, the existence of neutral mesons of mass about 280 m_e, and their mode of decay into two γ-rays, appears to have been put beyond doubt as a result of experiments by Panofsky, Aamodt and York (1950) on the capture of π^--particles in liquid hydrogen. This process leads to transmutations which may be represented by the following equations :—

$$\text{(i)}\ H^1+\pi^{-1} \rightarrow n^0+\pi^0 ; \quad \pi^0 \rightarrow 2h\nu\ (h\nu\sim70\ \text{MeV.}) ;$$

and

$$\text{(ii)}\ H^1+\pi^{-1} \rightarrow n^0+h\nu\ (h\nu\sim140\ \text{MeV.}) ;$$

π^0 representing a neutral meson.

The observed degree of homogeneity of the γ-rays from reaction (i), which is the more probable, indicates that the π^0-particles are emitted with small kinetic energy. This observation, and the known masses of the other particles involved in the transmutations, proves that the mass of the particle π^0 is only a few electron masses less than that of the charged π-particles ; viz. $m_{\pi^0}\sim280$ m_e.

These observations appear to have an important bearing on experiments carried out in the last few years with Wilson chambers by Cocconi *et al.* (1946), Fretter (1949), Chao (1949), Cocconi (1949), and Gregory *et al.* (1950). These authors have shown conclusively that the cascade showers of the soft component are commonly associated with penetrating showers of charged mesons.

2. SCOPE OF THE PRESENT EXPERIMENTS.

In Section I. of the present paper, observations are described which show that the detailed features of the spectrum of the γ-radiation accompanying the penetrating showers are consistent with the assumption that the quanta are created by the decay of neutral mesons of mass 295 ± 20 m_e ; and that the distribution in energy of the neutral particles is closely similar to that of the charged mesons of the " showers ".

In Section II., a method of determining the lifetime, τ_{π^0}, of the neutral mesons is described, and the results prove that $\tau_{\pi^0}<5\times10^{-14}$ sec. There is some indication that $\tau_{\pi^0}=3\times10^{-14}$ sec. ; and if this conclusion can be confirmed by further measurements, the observations would establish the independent existence of neutral mesons. Finally, the results show that the relative frequency of production of the neutral and charged mesons, $\mathcal{N}(\pi^0)/\mathcal{N}(\pi^\pm)=0{\cdot}45\pm0{\cdot}1$. It appears, therefore, that on the average, one neutral meson is produced for every two charged π-particles in the explosive disintegrations which give rise to the penetrating showers. This result, taken in conjunction with observations by Rossi on the relative amounts of energy carried by the " hard " and " soft " components of the cosmic radiation in the atmosphere, shows that the latter arises largely as a result of nuclear interactions, but indirectly through the production of neutral mesons ; and that at 70,000 ft., the fraction which need be attributed to other processes such as the decay of μ-mesons, " knock-on " electrons, etc. is negligible..

SECTION I.

Determinations of the Mass and the Distribution in Energy of Neutral Mesons.

3. THE γ-RAY SPECTRUM AT 70,000 FT.

During the past year observations have been made in this laboratory on photographic plates exposed at an altitude of 70,000 ft.; see Camerini *et al.* (1949). A number of these plates have been examined under the microscope with relatively high magnification, in order to detect pairs of electrons and other events involving only particles of charge e moving at relativistic velocities. Such events commonly escape observation when the plates are searched at lower magnification. By determining the energy of the pairs of electrons by the scattering method, in those favourable cases in which the tracks in the emulsion are of sufficient length, it has been possible to deduce the energy spectrum and direction of motion of the photons of the "soft" component at 70,000 ft.

TABLE I.

Energy region (MeV.)	Cell length (microns)	Probable error in energy (per cent)
10–100	100	19
100–200	200	28
200–400	250	32

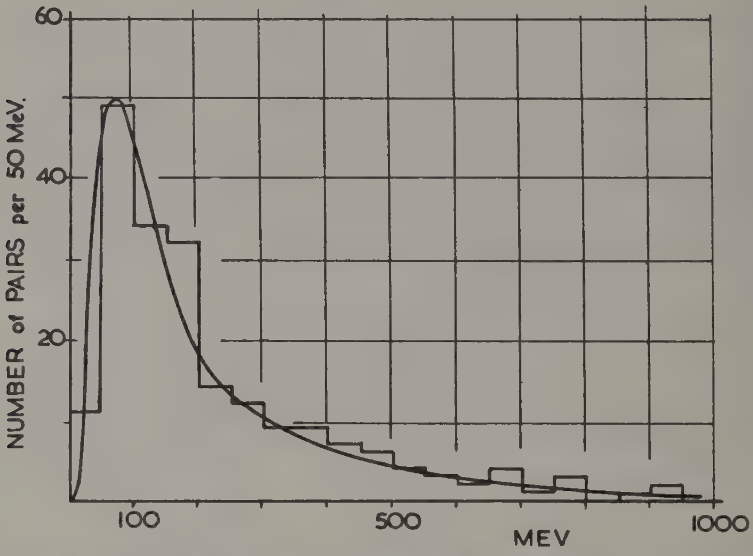
The above errors are given on the assumption that the track under consideration is just 1000 microns long.

The method employed in the determination of the energy of the electrons was similar to that described by Fowler (1950) and King (1950). Only those pairs were accepted for measurement of which the track of each component was longer than 1 mm. In the case of tracks of length less than 2000 μ, it is not possible to determine the energy of the corresponding particles unless it is less than 500 MeV. On the other hand, satisfactory estimates can be made up to 400 MeV. with tracks only 1000 μ long. Characteristic values of the probable errors of the measurements, for tracks 1000 μ long, are shown in Table I. The energy of a pair is usually substantially greater than that of one of its components. We may, therefore, say that the total energy, E, of a pair may be measured with a probable error of less than 30 per cent, for values of $E < 600$ MeV., if both tracks are longer than 1000 μ; and for $E < 1500$ MeV., if the tracks are longer than 2000 μ.

The results of measurements on pairs of very low energy require a small correction to allow for the influence of scattering on the escape of particles from the emulsion. Such corrections are, however, very small for $E > 20$ MeV., and they have not been applied in making the present measurements. The results are shown in fig. 1.

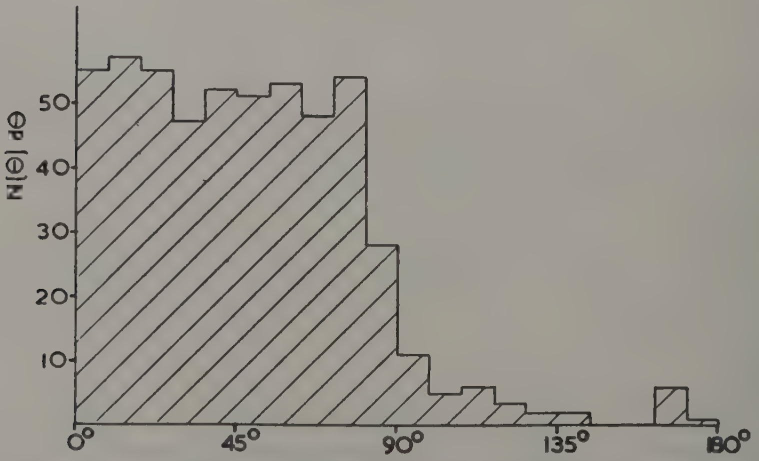
In determining the orientation of the pairs of electrons with respect to the vertical line through the plates during the exposure, it was considered sufficient to determine only the inclination, θ, of the projection of the "bisector" of the pair. For pairs of electrons produced by radiation with

Fig. 1.



Distribution in energy of the pairs of electrons observed at 70,000 ft. The full line shows the expected photon spectrum arising by the decay of neutral mesons of mass 280 m_e and with kinetic energies similar to those of charged π-particles emitted from stars.

Fig. 2.



Angular distribution of the directions of motion of photons at 70,000 ft.

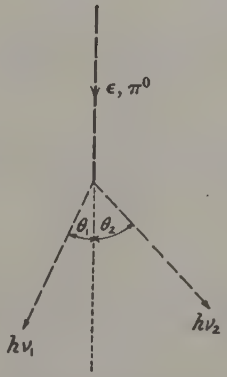
values of the quantum energy in the range under consideration, the "bisector" gives a measure of the direction of the parent γ-ray which is subject to a probable error of less than 0·2°.

The distribution in the values of θ, deduced from observations of 500 pairs is shown in fig. 2. It will be seen that the directions of motion of the quanta are distributed nearly isotropically in the range of angles from 0° to 90° with the direction vertically downwards, and that the intensity of any upward directed radiation is relatively very small. Charged mesons ejected in nuclear disintegrations in this exposure exhibit the same angular distribution; Davies, Franzinetti, and Perkins (1950).

4. Characteristics of the Radiation Produced by the Decay of Neutral Mesons.

Following the publication of the results of Bjorklund *et al.*, it occurred to us that if the γ-radiation in the atmosphere arises by the decay of neutral mesons—and if it is examined in conditions in which it has not been seriously modified by "bremsstrahlung" and other processes associated with the development of the cascade showers of the soft component— its spectrum should display certain features of an internal consistency, and should allow a determination to be made of the mass of the postulated neutral meson. The method depends on the following considerations :—

Fig. 3.



Suppose a neutral particle of mass m_0, moving with velocity βc, transforms into two quanta, of energy $h\nu_1$ and $h\nu_2$ emitted in directions making angles θ_1 and θ_2 with the line of motion of the parent particle—see fig. 3. The equations for the conservation of energy and of momemtum may be written :

$$\frac{h\nu_1}{c}\sin\theta_1=\frac{h\nu_2}{c}\sin\theta_2\,; \qquad \frac{h\nu_1}{c}\cos\theta_1+\frac{h\nu_2}{c}\cos\theta_2=m_0\beta c/\sqrt{(1-\beta^2)}\,;$$

$$h\nu_1+h\nu_2=m_0c^2/\sqrt{(1-\beta^2)}=\mathrm{B}m_0c^2=\epsilon.$$

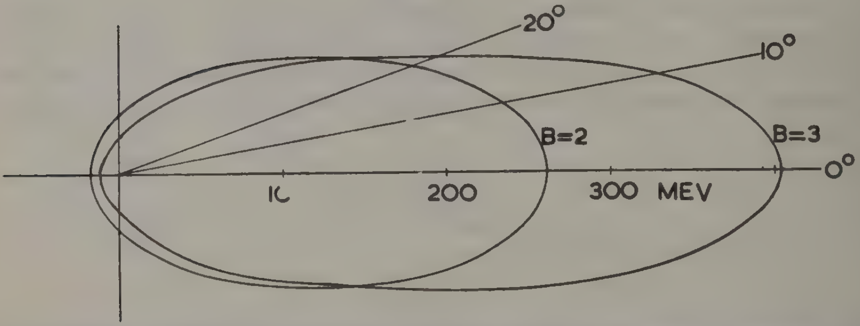
It is then easily shown that the energy of the quantum emitted at an angle θ is given by the well known formula for relativistic Doppler effect,

$$h\nu = \frac{m_0c^2}{2B(1-\beta\cos\theta)}. \quad \ldots \ldots \ldots \quad (1)$$

Characteristic curves showing the variation of $h\nu$ with θ, for different values of B and taking m_0c^2 equal to 140 MeV., are given in fig. 4. It follows from (1) that for a given value of β, the maximum and minimum values of $h\nu$ are given by the relations

$$h\nu\,(\max) = \frac{Bm_0c^2}{2}(1+\beta)\,; \qquad h\nu\,(\min) = \frac{Bm_0c^2}{2}(1-\beta)\,;$$

Fig. 4.



Polar diagram showing the variation of energy with the direction of emission of the photons resulting from the decay of neutral mesons. Calculations for two values of the total energy of the neutral mesons.

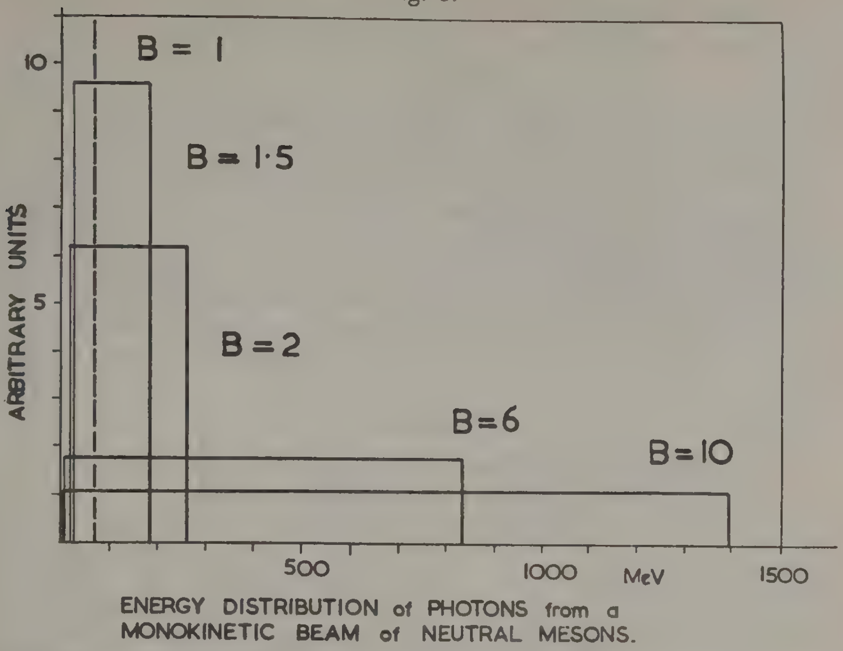
and that the greater the velocity of a homogeneous beam of neutral mesons, the greater the width of the spectrum of the emitted γ-radiation. Assuming that the γ-radiation is emitted isotropically in a coordinate system moving at a velocity equal to that of the mesons, the form of the spectrum of the γ-rays can be calculated—see Appendix—and can be shown to be given by the radiation relation

$$N(E)\,.\,dE = k\,.\,dE,$$

where $N(E)\,.\,dE$ is the number of emitted quanta of energy between E and $E+dE$. k is a constant for values of E lying between $h\nu$ (max) and $h\nu$ (min); and $k=0$ for values of E outside this interval. Characteristic spectra for different values of B, normalized to correspond to the same total number of quanta, are shown in fig. 5.

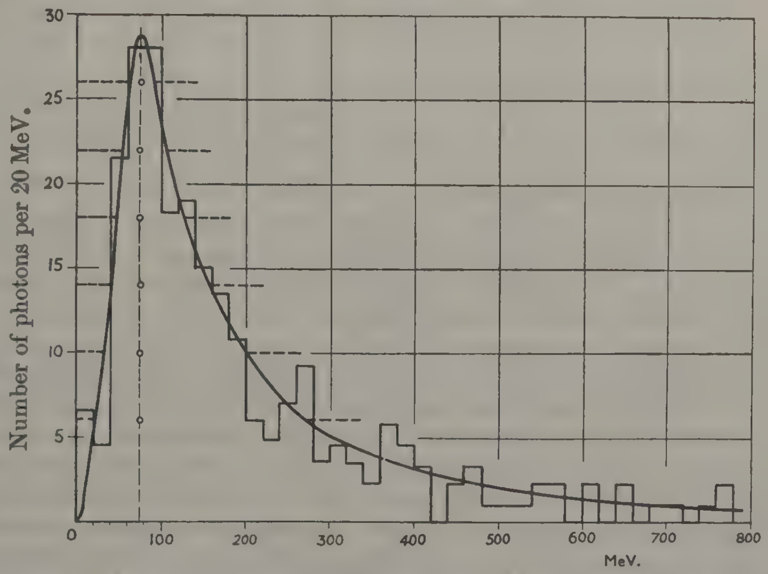
In the case of an inhomogeneous beam of neutral mesons, dispersed in velocity, the resulting γ-ray spectrum can be shown to have the features illustrated in fig. 6. There is a cusp in the spectral distribution if a finite number of neutral mesons are emitted with zero velocity. For any value

Fig. 5.



Distribution in energy of the photons produced by the decay of neutral mesons.

Fig. 6.



Spectrum of the γ-radiation at 70,000 ft. The full line is that giving the best fit with the experimental results, and the mass of the neutral particle has been deduced from it.

of the intensity below the maximum, there are two corresponding values of the quantum energy E_1 and E_2, and these can be shown to be related to the rest-mass m_0 of the parent neutral particles by the relation

$$\sqrt{(E_1E_2)}=m_0c^2/2. \qquad (2)$$

The proof of this theorem is included in the Appendix.

It follows from the above result that if most of the γ-radiation at 70,000 ft. is produced by the decay of neutral mesons—and if it is not seriously modified by the processes which lead to the formation of the cascade showers—there should be an inner consistency in the form of the spectrum corresponding to the relation displayed in equation (2). It appeared to us that the observation of such a consistency would give very powerful support for the view that the γ-radiation does in fact arise by a decay of neutral mesons and would allow an estimate to be made of their mass.

TABLE II.

Measurements of the Neutral Meson Mass from the Observed Energy Distribution of the Photons.

Weights (No. of pairs)	E_1	E_2	$\epsilon_0/2=\sqrt{(E_1E_2)}$	Weighted $\epsilon_0/2$
26	63	91	75·5	196
22	53	106	75·0	165
18	45	124	75·0	135
14	39	156	78·0	109
10	31	200	79·0	79
6	21	274	75·5	42

Total 96 Weighted mean for $\epsilon_0/2=75{\cdot}8$ MeV.

All energies are given in MeV. Mass of neutral particle $m_{\pi^0}=295\pm20m_e$.

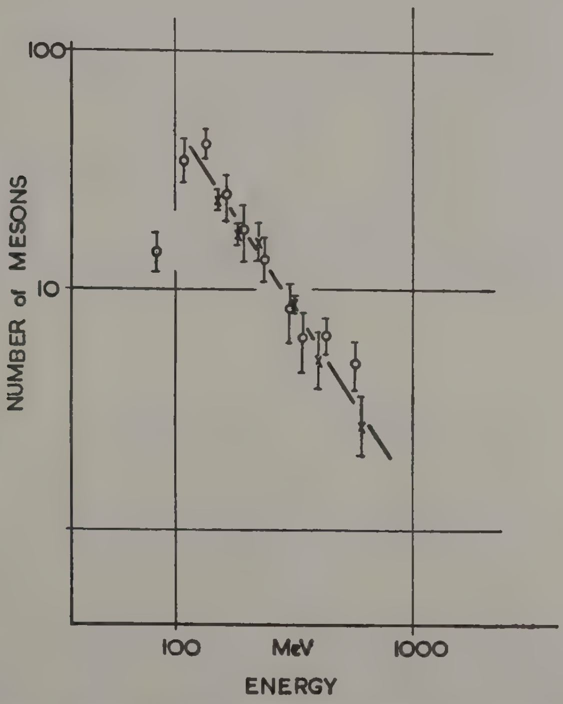
5. COMPARISON WITH EXPERIMENT ; THE MASS OF THE NEUTRAL MESON.

Fig. 6 shows the distribution in energy of the γ-radiation. It has been deduced from that of the pairs (see fig. 1) by assuming a variation with energy of the cross-section for pair production given by the quantum electro-dynamics (Heitler 1944). The full line in best accord with the distribution has been drawn, and intercepts made at various arbitrarily chosen intensities to determine values of E_1 and E_2. The corresponding values of $\sqrt{(E_1E_2)}$ are shown by circles. It will be seen from fig. 6 and Table II. that they show a remarkable degree of consistency, and indicate that the γ-radiation arises by the decay of neutral mesons of mass 295 ± 20 m_e. The results may be subject to a small systematic error arising from a variation with energy of the scattering constant k, defined by the equation $E=k/\bar{\alpha}$, where $\bar{\alpha}$ is the mean observed angle of scattering for a cell length of 100 μ. Such a variation may exist as a result of the effects of screening.

The stated errors in the estimated mass of the neutral meson correspond only to those due to statistical fluctuations in the numbers of measured pairs of electrons, and when the result was first obtained it was not clear to what extent it was influenced by the development of the cascade showers. The recent observations of Panofsky *et al.* (1950) suggest, however, that the " distortion " in the form of the spectrum is small, and this appears to be associated with the following features of the experiments.

Firstly, the plates were exposed near the main meson production layer in the atmosphere. At this altitude the probability that a given quantum will have produced a cascade, even in the case of quanta produced at the top of the atmosphere, is small. Secondly, if the cascade process starts in the plate assembly itself, it will commonly give rise to relatively narrow showers, made up of a number of particles moving nearly parallel to one another and close together. The development of such cascade showers has been observed in the present experiments, but they are rare, and any pairs associated with them have been excluded from the analysis.

Fig. 7.



The distribution in energy of the neutral mesons emitted from nuclear disintegrations. Spectrum of neutral mesons shown thus : × ; and of charged mesons, as observed by Camerini *et al.* (1950), shown thus : ○.

For these reasons we regard our observations as giving very strong evidence for the view that a large part of the γ-radiation in the atmosphere is due to the production and decay of neutral mesons, of mass $\sim 295 \pm 20 m_e$, which may be identified with the particles called for by the results of experiments by Bjorklund *et al.* (1949), and of Panofsky *et al.* (1950).

6. Distribution in Energy of the Neutral Mesons.

From the observed form of the γ-ray spectrum it is possible to deduce the distribution in energy of the parent neutral mesons and the result, determined by the methods described in the Appendix is shown in fig. 7. The corresponding distribution for the charged mesons, emerging from the nuclear explosions recorded in the same plates and determined by Camerini *et al.* (1950 a), is included in fig. 7, and it will be seen that the two "spectra" are similar in form. Alternatively, the spectrum of the γ-rays which would be produced by the decay of neutral mesons, of mass 280 m_e and with an energy distribution similar to that of the charged mesons, can be easily computed. Using the results of Camerini *et al.*, for the charged mesons, the result thus obtained is shown by the smooth line in fig. 1. It will be seen that the calculated form of the spectrum is indistinguishable from that observed, and added weight is thus given to the basic assumption regarding the origin of the γ-rays, and for the view that the parent neutral mesons are created with a distribution in energy closely similar to that of the charged particles.

Section II.

Observations to Establish the Independent Existence of Neutral Mesons, and to Measure their Lifetime and Frequency of Occurrence.

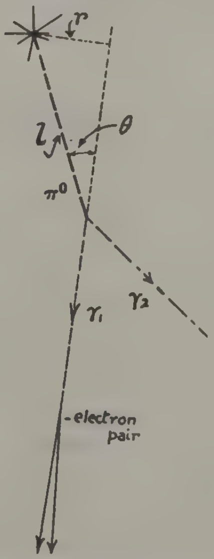
7. A Method of Determining the Lifetime.

The evidence for the existence of neutral mesons presented hitherto, whilst very strong, is circumstantial and depends on observing γ-radiation of which the detailed characteristics are consistent with an assumed origin by the decay of neutral mesons. Although the mass of evidence which has now been accumulated is so large that the conclusion to which it leads can hardly be doubted, it is nevertheless important to obtain an unambiguous proof of the independent existence of the particles. Further, it has been pointed out by Fukuda and Miyamoto (1949) that the determination of the lifetime of the neutral meson is of great importance for distinguishing between different theoretical treatments of the nature of the particle. The following considerations indicate that such evidence could be obtained by an application of the photographic method if the lifetime of the particles is greater than about 2×10^{-14} sec.

Suppose that a neutral meson emerges, with a total energy $\epsilon = Bm_0c^2$, from a nuclear explosion occurring in the emulsion; and that it decays into two photons, after traversing a distance l. We have seen that the two γ-rays will in general be emitted in directions inclined to that of the parent particle. Suppose, further, that one of the photons produces a pair of electrons in the emulsion within a distance of 1 or 2 mm. of its point of creation. The probability of such an occurrence is about 3 per cent for each emitted neutral meson; for the average length of path in the emulsion of a γ-ray before pair production is 46 mm. *, there are two photons produced by each neutral particle, and a considerable proportion of the events will escape observation because of the finite thickness of the emulsion.

We have seen that for neutral mesons with kinetic energies in the range of values with which we are concerned, the pair of electrons will move in directions inclined to one another at an angle of less than 0·2°.

Fig. 8.



Principle of the method for the determination of the lifetime of the neutral meson.

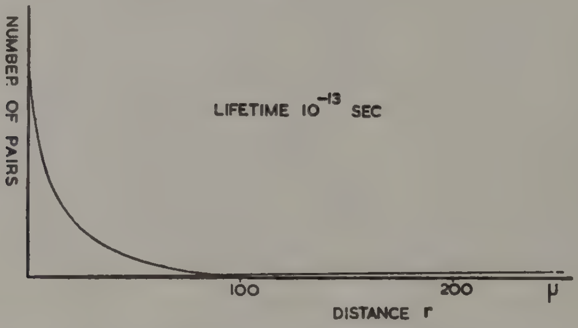
Apart from the momentum imparted to the nucleus involved in the formation of the pair (Bethe 1934), the bisector of the angle between the tracks of the two electrons therefore gives a measure of the direction of motion of the parent γ-ray which is subject to errors of the order of 0·2° (see fig. 8). It follows that if the mean line of the pair is produced backwards, it will not pass precisely through the disintegrating nucleus in which the neutral meson originated, but at a distance r from it. The greater the value of the lifetime, the greater the average distance l and the corresponding values of r.

---

* This value has been calculated by taking account of the observed energy spectrum of the γ-radiation, and the variation of the conversion length with quantum energy.

The value of r for any particular event will depend upon the distance travelled by the neutral meson before it decays, and the direction of emission of the photon relative to the line of motion of the parent particle. For a homogeneous beam of neutral mesons, moving with a given energy, the expected distribution in the values of r can be calculated by the methods outlined in the Appendix. From the results thus obtained, it is then possible to proceed to the general case in which the neutral mesons have a distribution in energy. Fig. 9 represents the results of such computations, assuming an energy distribution similar to that of the charged mesons deduced by Camerini *et al.* (1950 a), and a lifetime of the neutral mesons equal to 10^{-13} sec.

Fig. 9.



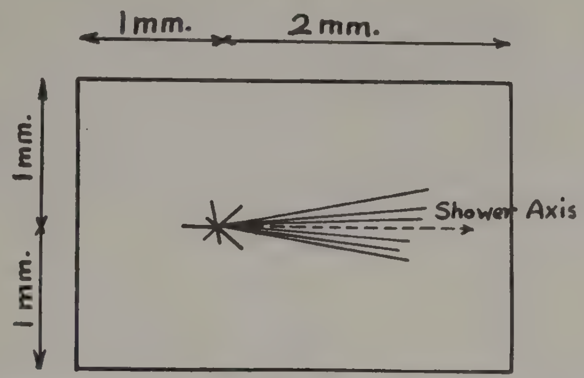
The calculated distribution in the values of the quantity r for an assumed lifetime of the neutral meson of 10^{-13} sec., together with the calculated form of the distribution due to unrelated pairs.

It is a particular feature of the method that in the case of mesons of high energy, $B \geqslant 2$, the calculated distributions are almost independent of the value of B, and of the distribution in energy of the neutral particles. This is due to the fact that although the emitted radiation tends to be closely collimated about the line of motion of the parent mesons as B increases—a factor which leads to a reduction in the values of r—the effect is compensated by the relativistic extension of the time-scale of the moving particles. This tends to make the neutral mesons survive over a longer length of path in the emulsion.

8. Experimental Results.

In order to apply the above method an extensive search has been made at high magnification of the region of the emulsion in the immediate neighbourhood of nuclear explosions occurring in the plates in which three or more " shower " particles are created ($n_s \geqslant 3$). All pairs of electrons observed in the volume of the emulsion defined in the manner illustrated in fig. 10 were recorded, and the corresponding values of r determined by observations on the direction of the projection of the mean line of the pair in relation to the star, and observations of " dip ". The methods are discussed in more detail in a later paragraph with special reference to the errors of measurement.

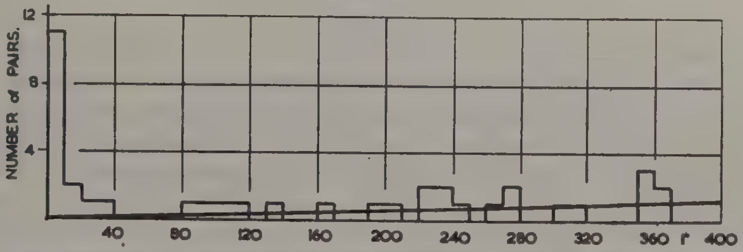
Fig. 10.



The area around a " shower ", which is scrutinized for " pairs ".

The observed distribution in the measured values of r is given in fig. 11, and displays a group of events corresponding to small values of r together with a pronounced general background. These features may be interpreted in the following manner:—We have seen that there is a random distribution of pairs of electrons in the plates due to the flux of γ-radiation, and these will sometimes, by chance, occur in the volume of emulsion round a star which is subject to scrutiny. The frequency of occurrence of these chance juxta-positions, which we may refer to as " unrelated " pairs, can be calculated from the observed numbers, per unit area of the plates, of the showers ($n_s \geqslant 3$), and of the pairs of electrons.

Fig. 11.

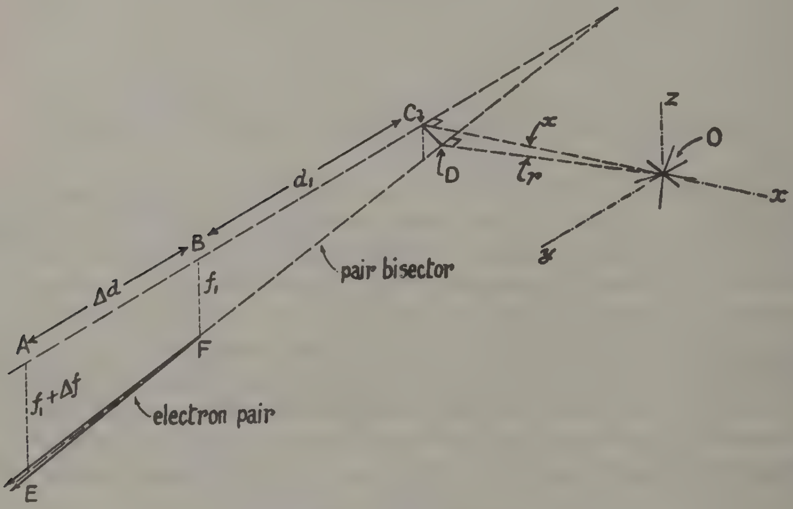


Observed distribution in the values of r for $r<400\ \mu$. The group of observations for small values of r, due to " related " pairs is clearly shown.

The calculated variation with r of the frequency of occurrence of these unrelated pairs is shown by the full line in fig. 11; and it will be seen to be in good accord with the observations. Thus in the interval in r from $40\ \mu$ to $200\ \mu$, the expected number of pairs is 22 ± 5 and the observed number 23. On the other hand, most of the fifteen pairs for which $r < 40\ \mu$ must be attributed to γ-radiation associated with the neighbouring explosion. Thus the expected number of unrelated pairs for this interval in the value of r is only 0·2 compared with the 15 observed. The observations therefore prove that γ-radiation is produced, by some mechanism, as a result of nuclear interactions leading to the production of shower particles of multiplicity $n_s \geqslant 3$.

The observations on the related pairs prove that the lifetime of the neutral mesons is less than 5×10^{-14} sec., and they appear to indicate that $\tau_{\pi^0} = 3 \times 10^{-14}$ sec. It has already been emphasized that such a result, apart from its interest in establishing the decay-constant for the neutral π-particles, would also provide a direct proof for their independent existence. It is therefore important to examine what weight can be put on the observations; and, in particular, whether they can be interpreted, alternatively, as due to errors of measurement, the lifetime of the π^0-particles being much shorter than can be established by this method.

Fig. 12.



Determination of the quantity r.

9. ERRORS OF MEASUREMENT IN THE DETERMINATION OF THE QUANTITY r.

The quantity r which is taken as a measure of the relationship of a pair to a particular nuclear explosion, is illustrated in fig. 12. It is the perpendicular distance from the centre of the star to the bisector of the lines of motion of the two electrons into which the γ-ray transforms. To determine r, the perpendicular distance, x, from the star to the vertical plane containing the bisector of the electron-pair is first measured. By "vertical" is meant here the direction normal to the plane of the emulsion which is taken as "horizontal". x is represented by the line OC in fig. 12, and is determined by turning the plate until the bisector is parallel to one of the stage motions, Y, and coincident in the field of view with a line in the eye-piece graticule. The graticule is also provided with a scale perpendicular to the line. After moving the stage parallel to the Y axis, and refocusing, the star is eventually brought into the field of view. Its distance from the line can then be measured by means of the eye-piece scale.

The above method of determining x is subject to errors which depend on the distance BC$=d$ in fig. 12, and on the precision with which the bisector of the pair can be defined. Studies of scattering of fast particles (Davies *et al.* 1949) indicate that the alignment with a rectilinear track can be made with an error of less than 0·1°. In the case of the pairs produced by high energy photons, the two tracks are not appreciably scattered in the first 100 μ of the trajectory, and the angle between them, δ, is small; for 100 MeV. γ-rays, $\delta \sim 0{\cdot}2°$, and for 400MeV., $\delta \sim 0{\cdot}05°$. For such energetic radiation, which constitutes most of the spectrum (see fig. 1) the definition of the bisector can be made with a precision of the order of 0·25° and the corresponding error in x can be written $\delta x \sim \frac{d}{200}$, δx and d both being measured in microns.

The second quantity necessary for the determination of r is obtained from observations on the angle of "dip" of the bisector of the pair. In fig. 12, AC is the projection of the bisector on a plane, parallel to the surface of the emulsion, which contains the centre of the star, O, and f is the depth below this plane of the point of origin of the pair, F. At a distance Δd along the bisector, at the point E, the depth is $f+\Delta f$. It is then easily shown that

$$z^2 = \mathrm{CD}^2 = (d\,.\,\Delta f - f\,.\,\Delta d)^2/(\Delta d^2 + \Delta f^2),$$

and that

$$r^2 = x^2 + z^2.$$

In order to determine the probable errors resulting from the errors of measurement of the depth, numerous individual observations have been made on a number of events, and the standard deviations in the results deduced empirically. Similar observations have been made on protons of great energy which are observed to produce disintegrations, and which are known from scattering measurements to suffer an average change in direction along the trajectory of less than 0·01° per 100 μ. Values of "r" can thus be deduced for tracks which are known to pass through the centre of the star, and of which the departure from rectilinearity is negligible.

These observations indicate that the probable errors in the determination of the angle of dip are about 0·2° and that the uncertainty in the true values of r is contributed to almost equally by the two measurements, of x and of z. According to the particular characteristics of the individual events, one or other of the two measurements may be the more important. Taking both measurements into account, the probable error in the determination of the direction of the bisector is estimated to be 0·3°.

The important question remains of the precision with which the bisector gives a measure of the line of motion of the γ-ray. In general the lines of motion of the two electrons are not equally inclined to that of the γ-ray which produces them. Further, the nucleus involved in the creation of the pair receives an impulse of order of magnitude m_ec. (Bethe 1934, Modesitt and Koch 1950). At the energies with which we are concerned, the probable errors resulting from these effects can hardly exceed 0·2°. We thus obtain 0·4° as the final figure for the probable error in the determination of the line of motion of the photon. It follows that for errors in the measurement of the perpendicular distance from the star to this line, we may write $\delta r \approx d/150$.

TABLE III.

Individual r measurements and their errors.

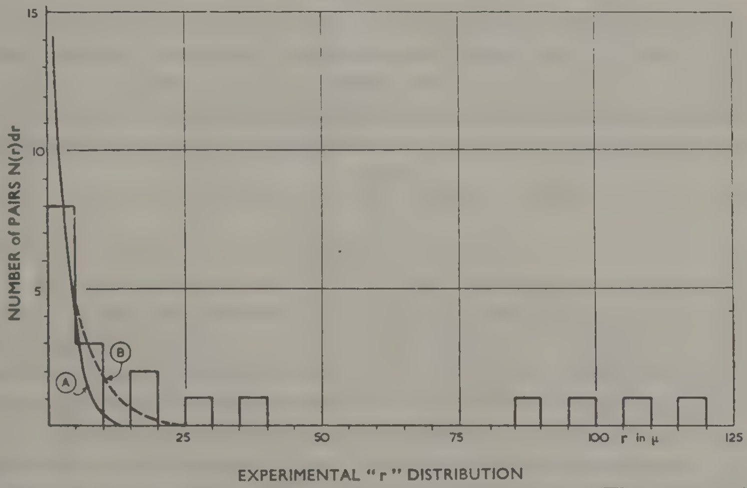
Pair	Star type	x	d	dip$=\Delta f/\Delta d$	f	r
1	7+6N	0	0	—	—	0
2	6+4p	0·5±0·5	73	0·050±0·013	2·4±0·5	1·5±1·5
3	3+5p	0·0±1·0	123	0·096±0·010	9·5±0·7	2·0±2·0
4	12+3p	2·0±2·0	497	0·120±0·010	60·0±8·0	2·0±4·0
5	9+7p	1·0±1·0	722	0·122±0·004	91·5±0·8	3·5±3·5
6	6+7p	3·5±1·0	1035	1·011±0·020	132·9±0·4	3·5±2·5
7	7+12p	1·0±1·0	378	0·094±0·004	31·8±0·5	3·5±2·0
8	21+18p	3·5±2·0	660	0·172±0·003	113·6±0·4	4·5±2·0
9	3+26p	3·0±2·0	1027	0·158±0·002	157·5±0·3	5·0±3·5
10	7+6N	6·0±1·0	226	0·337±0·013	78·6±0·5	7·0±2·0
11	3+26p	2·0±2·0	536	0·042±0·016	27·3±0·4	7·5±5·5
12	13+5p	6·0±3·0	1993	·0008±·0008	17·4±1·2	17·0±3·0
13	17+6p	7·0±1·0	395	0·320±0·006	145·0±1·1	17·5±3·0
14	14+3N	10·0±3·0	1080	0·049±0·003	81·6±0·8	29·0±4·0
15	11+4N	23·0±5·0	745	0·170±0·004	94·6±0·7	38·0±4·5

All the errors are expressed in terms of the standard deviation.
x, d, f and r are all measured in microns.

Details of the observations on the fifteen related pairs are given in Table III. The limits of error given in the last column are those deduced from the standard deviations associated with the inconsistence in a large number of measurements on the same event. They do not, therefore, nclude contributions due to a deviation between the direction of the γ-ray and the bisector of the pair.

A histogram showing the values of r on a more extended scale has been reproduced in fig. 13. Included in the figure are curves indicating the distributions that would be expected for standard deviations in the errors of measurement, arising from all sources, of 0·25 and 0·5°, the lifetime of the particles being assumed to be short compared with 10^{-14} sec. It will be seen that there are only two or three pairs for which the values of r are substantially greater than those expected from a standard deviation of 0·5°. Reference to Table III. shows, however, that the values of r are many times greater, for these four events, than the standard deviations. It is very improbable that four such events should be found among a sample of fifteen related pairs, if the γ-rays diverge, in effect, from the nuclear explosion; and we have seen that the number of expected unrelated pairs for $r<40\ \mu$ is only 0·2.

Fig. 13.



Distribution in the observed values of r for related pairs. The curves (A) and (B) show the expected distributions due to probable errors of measurement in the determination of the line of motion of the γ-ray of 0·25° and 0·5° respectively, the lifetime of the neutral mesons being assumed indefinitely short.

The number of events observed is not great enough, however, to allow us to put any weight on the indications of a finite lifetime given by the experiments, and we only note that if the apparent features can be confirmed by further observations of greater statistical weight, it may be possible to establish the magnitude of the lifetime if it is equal to or greater than 2×10^{-14} sec. The conclusion that the lifetime is less than 5×10^{-14} sec. appears to be established. It will be remarkable if the lifetime of the neutral particles is found to be just sufficiently great to be determined by this method, the only one available with the resources at present at our disposal.

10. THE INTENSITY OF THE NEUTRAL MESONS.

In order to determine the frequency of occurrence of the neutral mesons in comparison with the charged mesons, the total length, L, of the tracks of the shower particles in the scrutinized area round each star has been determined. Recent experiments indicate that about 80 per cent of these shower particles are π-mesons, and the contribution of fast protons can thus be excluded. Suppose that, on the average, the emission of each charged π-particle is accompanied by ϕ neutral mesons :

$$\mathcal{N}(\pi^0)/\mathcal{N}(\pi^{\pm})=\phi.$$

The length of path in the emulsion of the γ-rays formed by the decay of the neutral mesons is then given, without serious error, by

$$\frac{4}{5}\mathrm{L}\,.\,2\,.\,\phi=1\cdot6\mathrm{L}\phi.$$

The observation show that in this length of path, 15 pairs of electrons were created. Taking into account the small change with quantum energy of the " conversion-length ", the average length of path of a γ-ray in the emulsion before it produces a pair was calculated to be 4·6 cm., for the particular spectral distribution with which we are concerned (Heitler 1944). We may therefore write $\frac{1\cdot6\mathrm{L}\phi}{15}=4\cdot6$; with L=95 cm. it follows that

$$\phi=0\cdot45\pm0\cdot10.$$

The above result suggests that in the production of showers of charged particles, for every two charged mesons, one neutral particle is created. This conclusion receives support from a recent study by Camerini, Fowler, Lock and Muirhead (1950 b) in this laboratory, of the energy balance in the nuclear processes leading to the production of showers. These authors have compared the energy of the primary particles with that of the mesons and nucleons ejected in the disintegration. They have thus been able to show that there is an apparent disappearance of energy which can be accounted for if neutral π-particles are created. A similar result has been obtained by Green and Fretter (1950).

11. CONCLUSIONS.

We regard our experiments as proving that in the explosive disintegration of nuclei, neutral mesons are created in numbers equal to about half those of the charged π-particles. The mass of these neutral mesons is 295 ± 20 m_e, and they disintegrate into two γ-rays with a lifetime less than 5×10^{-14} sec. ; they may therefore be identified with the neutral π-mesons discovered in Berkeley. It is well known from the work of Steinberger (1949), and of Fukuda and Miyamoto (1949), that the mode of decay into two γ-rays indicates that the neutral meson is of spin zero.

The distribution in energy of the neutral mesons is closely similar to that of the charged mesons, and they arise in nuclear processes of the same type. They may therefore be regarded as having properties similar to those of neutral " heavy quanta " of Yukawa's theory.

The observed frequency of occurrence of the neutral mesons allows a simple explanation to be given of an observation by Rossi (1948) that the flows of energy represented by the " hard " and " soft " components of the cosmic radiation are of closely similar magnitude; an observation which appeared as an empirical fact at the time but which, as Rossi suggested, might have an important, deeper significance. Our results confirm the suggestion that both the " hard " and " soft " components arise in the nuclear disintegrations produced by protons and heavier particles of great energy, which are accompanied by the creation of charged and neutral π-particles. It appears to be unnecessary to attribute any of the soft component to " primary " electrons.

The lifetime of the π°-particles is certainly less than 5×10^{-14} sec.; and it will be difficult to determine it directly, and to establish the independent existence of the particles, with the methods at our disposal. A determination of the lifetime may, however, be possible by the present methods if it lies in the interval from 2×10^{-14} sec. to 5×10^{-14} sec.

ACKNOWLEDGMENTS.

We wish to take this opportunity of thanking Professor C. F. Powell, F.R.S. for his continuous interest and encouragement in the course of this work. We are indebted to Dr. H. K. Heitler who prepared the diagrams, to Mr. R. A. Gattiker and Miss P. M. Dyer who carried out the photography and to Messrs. W. O. Lock and H. Muirhead for processing the emulsions. Without the work of Mrs. D. M. Ford, Mrs. J. Cowie, Miss J. Jones and Miss. M. Jones, who performed the arduous scrutiny of the plates with oil-immersion objectives, this investigation would not have been possible.

The authors were supported by grants from the Swedish Atomic Committee, the Department of Scientific and Industrial Research, and the Medical Research Council respectively.

APPENDIX.

For completeness, we outline the main features of the calculations. We have seen, that the energy of a quantum emitted at an angle θ with respect to the line of motion of a neutral meson of velocity $v=\beta c$, total energy $\mathrm{B}m_0c^2=\epsilon$ is given by the equation

$$h\nu=\mathrm{E}=\frac{m_0c^2}{2\mathrm{B}(1-\beta\cos\theta)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

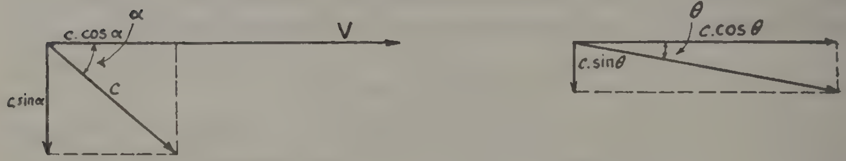
(a) *Angular Distribution of the Radiation.*

The quanta produced by the spontaneous decay of a homogeneous beam of neutral mesons may be assumed to be emitted isotropically in a coordinate system—the C-system—moving with the same velocity as the particles. The motions in this system are then transformed to the laboratory system—the L-system.

Suppose the angle of ejection of a quantum in the C-system, relative to the direction of motion of the meson, is α. The components of the velocity, c, of the quantum, parallel and perpendicular to the direction of motion of the particle, are $c \,.\, \cos\alpha$ and $c \,.\, \sin\alpha$, respectively (fig. 14). Let u_x and u_y be the corresponding components in the L-system. By the Einstein theorem for the addition of velocities we obtain

$$u_x = \frac{c\cos\alpha + v}{\left(1 - \dfrac{v\cos\alpha}{c}\right)}; \qquad u_y = \frac{c\sin\alpha}{\mathrm{B}\left(1 - \dfrac{v\cos\alpha}{c}\right)},$$

Fig. 14.



where $\mathrm{B} = 1/\sqrt{(1-\beta^2)}$. It follows that

$$\tan\theta = \frac{u_x}{u_y} = \frac{\sin\alpha}{\mathrm{B}(\cos\alpha + \beta)}.$$

Writing $\sin\xi = \sqrt{(1-\beta^2)}$ and $\cos\xi = \beta$, it follows that :

$$\tan\theta = \frac{\sin\alpha\sin\xi}{\cos\alpha + \cos\xi}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

From the identity $\sec^2\theta = \tan^2\theta + 1$, we then obtain the useful relations

$$\cos\theta = \frac{\cos\xi + \cos\alpha}{1 + \cos\xi\cos\alpha}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$\sin\theta = \frac{\sin\xi\sin\alpha}{1 + \cos\xi\cos\alpha}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and from (2), (3) and (4)

$$\frac{d\alpha}{d\theta} = \frac{(\cos\xi + \cos\alpha)^2 \,.\, \sec^2\theta}{\sin\xi(1 + \cos\xi \,.\, \cos\alpha)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Let the intensity of the radiation in the L-system, measured as the number of quanta per unit solid angle, be I(θ); and in the C-system, I(α). We can then write

$$I(\theta)\,.\,\sin\theta\,.\,d\theta = I(\alpha)\,.\,\sin\alpha\,.\,d\alpha = K\,.\,\sin\alpha\,.\,d\alpha.$$

Putting K=1 for convenience, we obtain

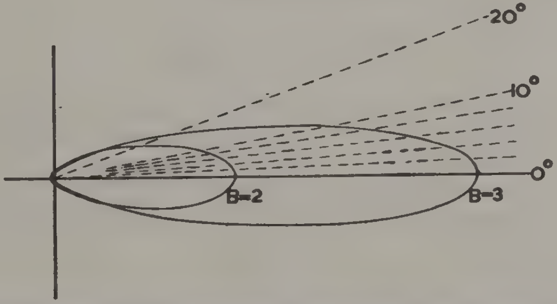
$$I(\theta)=\frac{\sin\alpha}{\sin\theta}\cdot\frac{d\alpha}{d\theta};$$

which reduces to

$$I(\theta)=1/B^2(1-\beta\cos\theta)^2 \quad . \;.\;.\;.\;.\;.\;.\;.\;. \quad (6)$$

The form of the distribution, for two values of B, is shown as a polar diagram in fig. 15.

Fig. 15.



Polar diagram showing the distribution in angle of the intensity of the γ-rays produced by the decay of homogeneous beams of neutral mesons of different energies.

(b) *Distribution in energy of the gamma-rays.*

We may write $N(E)\,.\,dE=\frac{1}{2}I(\theta)\,.\,\sin\theta\,.\,d\theta$. From (1) and (6) we obtain

$$\frac{d\theta}{dE}=-\frac{2(1-\beta\cos\theta)^2}{\beta\epsilon(1-\beta^2)\sin\theta}.$$

and, thence,

$$N(E)=1/\beta\epsilon=\text{constant}. \quad . \;.\;.\;.\;.\;.\;.\;. \quad (7)$$

The distribution extends from

$$E_1=(1-\beta)\epsilon/2 \qquad \text{to} \qquad E_2=(1+\beta)\epsilon/2. \quad . \;.\;.\;. \quad (8)$$

(c) *Properties of the γ-ray spectrum.*

We now consider an inhomogeneous beam of neutral mesons distributed in energy, the number in the interval from ϵ to $\epsilon+d\epsilon$ being $F(\epsilon)\,.\,d\epsilon$. The γ-ray intensity due to the neutral mesons in this energy interval is given, according to (7) by the equation

$$\frac{F(\epsilon)\,.\,d\epsilon}{\beta\epsilon}=\frac{F(\epsilon)\,.\,d\epsilon}{\sqrt{(\epsilon^2-\epsilon_0^2)}} \quad . \;.\;.\;.\;.\;.\;.\;. \quad (9)$$

between the limits

$$E_1=\tfrac{1}{2}\{\epsilon-\sqrt{(\epsilon^2-\epsilon_0^2)}\} \text{ and } E_2=\tfrac{1}{2}\{\epsilon+\sqrt{(\epsilon^2-\epsilon_0^2)}\}, \quad . \quad . \quad (10)$$

where ϵ_0 is the rest-energy of the neutral meson. At a γ-ray energy, E, the intensity is obtained by integrating equation (9) between the limits $\epsilon=E+\epsilon_0^2/4E$ and infinity; for every neutral meson with energy above ϵ contributes to the γ-ray intensity at the energy E. This holds both for energies $E_1\leqslant\epsilon_0/2$ and $E_2\geqslant\epsilon_0/2$, because, when solving (10) for ϵ, we obtain in both cases

$$\epsilon=E_{1,2}+\epsilon_0^2/4E_{1,2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Thus the γ-ray spectrum is given by

$$N(E)=\int_{\epsilon}^{\infty}\frac{F(\epsilon)\,d\epsilon}{\sqrt{(\epsilon^2-\epsilon_0^2)}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

By differentiating (12) we obtain

$$F(E+\epsilon_0^2/4E)=E\left|\frac{dN}{dE}\right| \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

a relation which allows the distribution in energy of the mesons to be derived from the observed spectrum of the γ-radiation. At the quantum energy $E=\epsilon_0/2$, the intensity of the γ-radiation has a maximum value. It is clear from (13) that the slope, dN/dE, at this point is zero only when $F(\epsilon_0)=0$; *i. e.*, when there are no mesons of zero velocity in the beam.

(d) *The Rest Energy of the Neutral Meson.*

It follows from (12), which holds both for $E_1\leqslant\epsilon_0/2$ and $E_2\geqslant\epsilon_0/2$, that the intensity of the γ-radiation is the same for E_1 and E_2 when according to (11)

$$E_1+\epsilon_0^2/4E_1=E_2+\epsilon_0^2/4E_2$$

which gives

$$\epsilon_0^2=4E_1E_2$$

or

$$\epsilon_0=2\sqrt{(E_1E_2)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

This relation, which permits the derivation of the rest energy, ϵ_0, from the observed γ-ray spectrum, holds generally, being a result of the Doppler effect. It is therefore independent of the form of the " spectrum " of the neutral mesons.

(e) *Derivation of the r Distributions.*

Consider a meson emerging from a star (fig. 8), travelling a distance l before decaying. Assume that it then emits one of its γ-rays in the direction making angle θ with the direction of motion of the neutral meson, so that the orthogonal distance from the star to the line of motion of the γ-ray is r. It follows that

$$r=l\,.\,\sin\theta\,;$$

and, as we have $N(\theta) \,.\, d\theta = I(\theta) \,.\, \sin\theta \,.\, d\theta$, we may write

$$N(r) = N(\theta)\frac{d\theta}{dr} = I(\theta) \,.\, \tan\theta/l.$$

Now $l = c\beta Bt$, where t is the time interval between the instants of creation and the decay of the neutral meson in the C-system. We assume that the meson decays according to the usual laws of radioactive decay, and then find that the number decaying between times t and $t+dt$ is

$$N(t)\,dt = \frac{N_0}{\tau_0}\exp(-t/\tau_0)\,dt,$$

τ_0 being the proper mean lifetime of the meson. Hence,

$$N(l)\,dl = \frac{N_0}{c\beta B\tau_0} \,.\, \exp(-l/c\beta B\tau_0) \,.\, dl\,;$$

and since

$$l = r/\sin\theta,$$

$$N(l)\,dl = \frac{N_0}{c\beta B\tau_0} \cdot \frac{r}{\tan\theta\sin\theta}\exp(-r/c\beta B\tau_0\sin\theta)\,d\theta\,;$$

and

$$N(r) = N_0/c\beta B\tau_0 \,.\, \int_0^\pi I(\theta)\exp(-r/c\beta B\tau_0\sin\theta)\,d\theta.$$

We have performed this integration graphically for several values of the energy, and found that the calculated distributions in the values of r are nearly exponential for values greater than 10 μ. For values $B \geqslant 2$ the curves are very nearly identical.

The final distribution is obtained by adding the curves for different values of B, each appropriately weighted according to the energy spectrum of the neutral mesons.

(f) *Background Pairs.*

The probability of finding an unrelated pair of which the bisector passes the star at a distance lying between r and $r+dr$ may be calculated from the known angular distribution and frequency of occurrence of such pairs in the plates.

Our observations show that γ-rays are isotropically distributed in the lower hemisphere (fig. 2). As a result of the geometry of the scanned area (fig. 10), two-thirds of the unrelated pairs point towards the star. If $N(r)\,dr$ is the distribution in r formed by these pairs, we may write

$$N(r) = \tfrac{1}{4}\int_r^R I(\theta) \,.\, \sin\theta \,.\, \frac{d\theta}{dr} \,.\, N(l) \,.\, dl,$$

where l is the distance between the star and the origin of the pair, θ is the angle between the bisector and the line joining the origin of the pair with the star, and $N(l)\,dl$ is the number of pairs of which the origin lies in the spherical shell of radii l and $l+dl$. Thus $N(l)\, dl=\rho 4\pi l^2\, dl$ where ρ is the observed number of pairs per unit volume. The angular distribution $I(\theta)$ is constant. R is the effective radius of the scanned area. Using the relation $r=l\,.\,\sin\theta$ we obtain

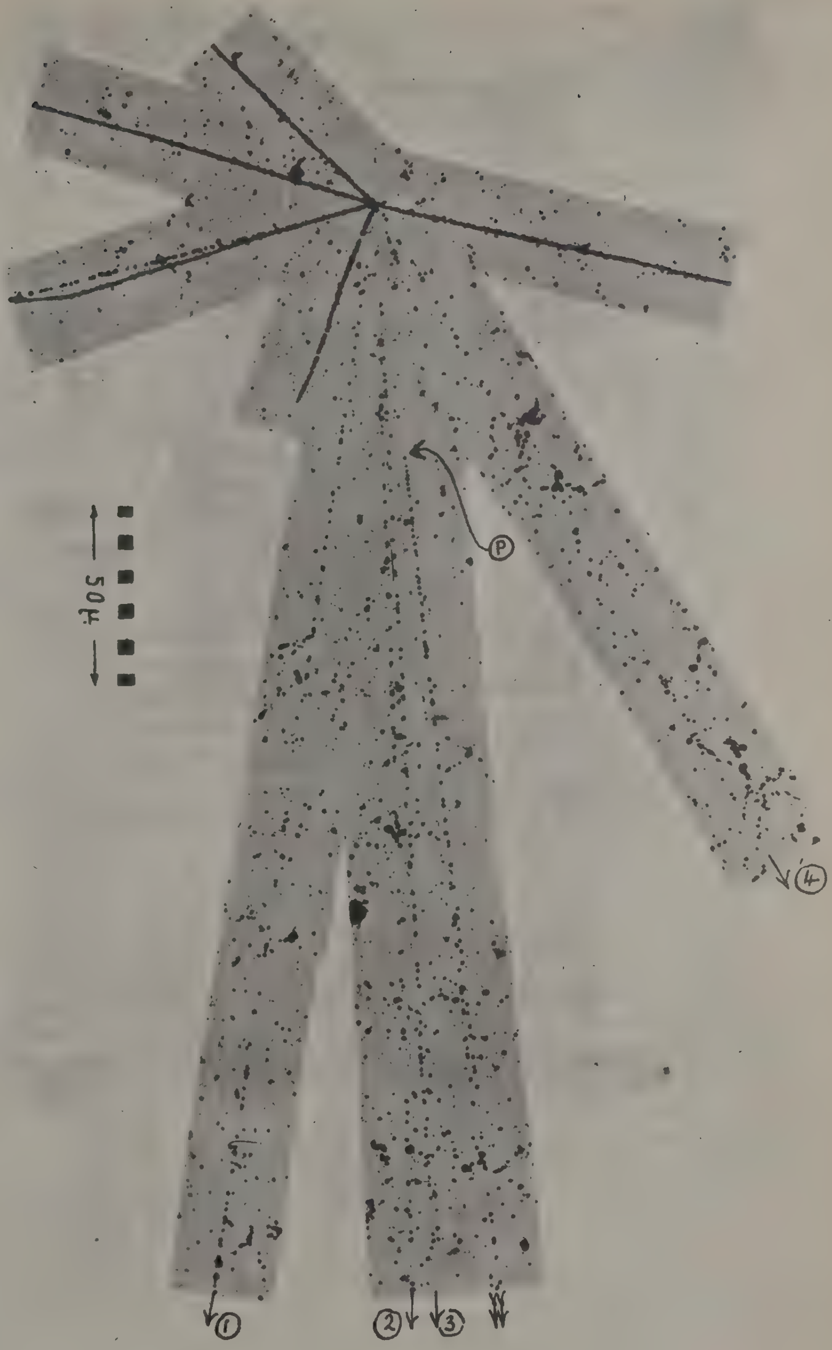
$$N(r)=\text{constant}\times\int_r^R l\,.\,\tan\theta\, d\theta$$

$$=\text{constant}\times r\sqrt{(R^2-r^2)}.$$

For $r\leqslant 600\ \mu$, we have $N(r)\sim\text{const}\,.\,R\,.\,r$, a relation which is true to within 10 per cent. The constant is then deduced from the observed density of " background " pairs and the known value of R.

REFERENCES.

BETHE, 1934, *Proc. Camb. Phil. Soc.*, **30**, 524.
BJORKLUND, CRANDALL, MOYER and YORK, 1950, *Phys. Rev.*, **77**, 213.
BRADT, KAPLON, and PETERS, 1950, *Helv. Phys. Acta.*, **23**, 24.
CAMERINI, COOR, DAVIES, FOWLER, LOCK, MUIRHEAD and TOBIN, 1949, *Phil. Mag.*, **40**, 1073.
CAMERINI, FOWLER, LOCK and MUIRHEAD, 1950 a, *Phil. Mag.*, **41**, 413 ; 1950 b, *Ibid.* (in course of publication).
COCCONI, LOVERDO and TONGIORGI, 1946, *Phys. Rev.*, **70**, 852.
COCCONI, 1949, *Rev. Mod. Phys.*, **21**, 26.
CHAO, 1949, *Phys. Rev.*, **75**, 581.
DAVIES, FRANZINETTI and PERKINS, 1950, *Phil. Mag.* (in course of publication).
DAVIES, LOCK and MUIRHEAD, 1949, *Phil. Mag.*, **40**, 1250.
FOWLER, 1950, *Phil. Mag.*, **41**, 169.
FRETTER, 1949, *Phys. Rev.*, **76**, 511.
FUKUDA and MIYAMOTO, 1949 a, *Prog. Theor. Phys.*, **4**, III, 347 ; 1949 b, *Ibid.*, **4**, III, 391, 392.
GREEN and FRETTER, 1950 (Private communication).
GREGORY, ROSSI and TINLOT, 1950, *Phys. Rev.*, **77**, 299.
HEITLER, 1944, *Quantum Theory of Radiation*, 2nd Edition (Oxford : University Press).
KING, 1950, *Nature, Lond.*, **165**, 526.
MODESITT and KOCH, 1950, *Phys. Rev.*, **77**, 175.
PANOFSKY, AAMODT and YORK, 1950 (Private communication).
PICCIONI, 1950, *Phys. Rev.*, **77**, 1.
ROSSI, 1948, *Rev. Mod. Phys.*, **20**, 537 ; 1949, *Ibid.*, **21**, 104.
STEINBERGER, 1949, *Phys. Rev.*, **76**, 1180.



A large shower of mesons (6+4p) showing the origin of a related pair very close to the star. The pair has been marked by an arrow in the microphotograph.

Observer: Miss E. James.

To face page 724

LXIV. CORRESPONDENCE

Gap Measurement as a Method of Analysing Cosmic Ray Stars in Photographic Emulsions.

By P. E. Hodgson *.

[Received May 31, 1950.]

Owing to the recent increase in sensitivity of photographic emulsions for nuclear research, tracks of particles of moderate energy are clogged, that is, the grains coalesce to form continuous blocks. Since grain counting is a well-known method of identifying and estimating the energy of such particles, it is desirable to extend the method to such tracks.

While in regions of moderate clogging some simple convention, such as one micron of solid track is equivalent to three grains, is satisfactory, this is highly artificial where the clogging is severe. In such regions it is natural to measure the total gap length per unit length, which is defined as the gap density. As grain counting is easier and more rapid than gap measurement, it is desirable to use it to the highest possible grain densities, which in practice are found to be in the region of 1·5 grains per micron, which corresponds to a gap density of about 0·1. Beyond this region gap measurement must be used.

In order to relate the gap density to the energy loss, gap measurements versus residual range were made on the tracks of a number of particles of different mass and charge ending in the emulsion of an Ilford G5 plate exposed on the Jungfraujoch. Figs. 1 and 2 show the discrimination obtained by means of graphs of total gap length against residual range. If the track is longer than 400 microns, gap measurement gives almost certain identification, except between protons, deuterons and tritons, and between π- and μ-mesons. The two τ-mesons are those recently reported (Harding 1950).

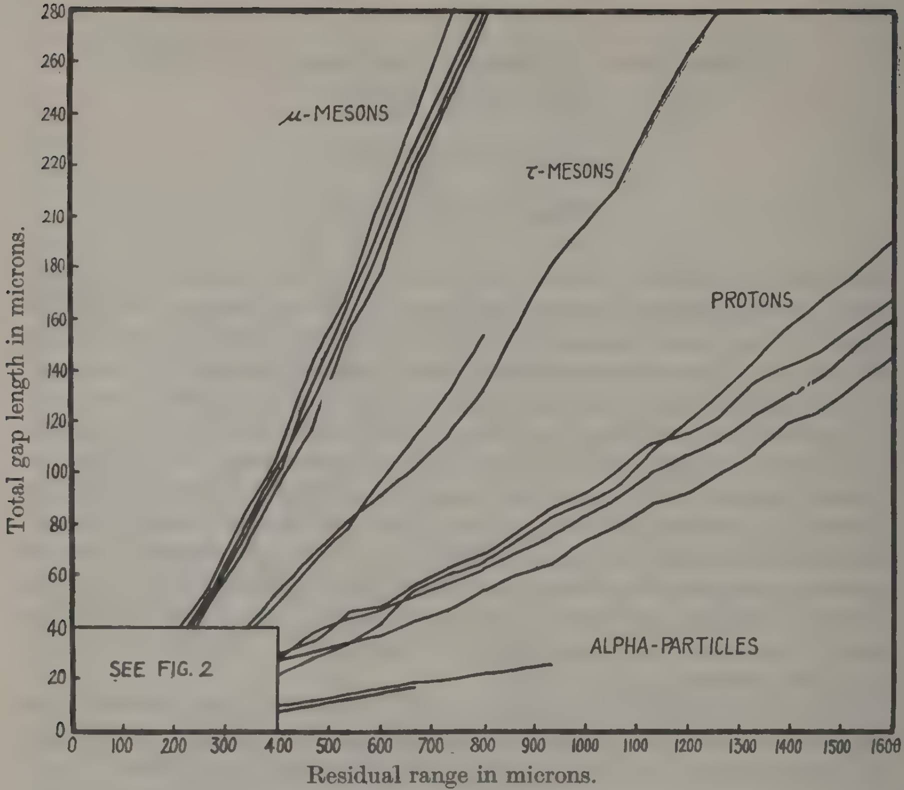
The energy losses at points along the tracks were calculated from the known range–energy curves (Bradner *et al.* 1950), and hence a graph of gap density versus energy loss was constructed. These results are plotted on fig. 3, together with the standard deviations of measurements along 130 micron samples of track. In practice more than 130 microns of track are often available, and so the accuracy is higher than indicated by the figure. It is found that, above energy losses of 9 KeV./μ, the gap density and energy loss satisfy the empirical formula

$$\frac{dG}{dR}=0{\cdot}53\left(\frac{dE}{dR}\right)^{-0\cdot7},$$

---

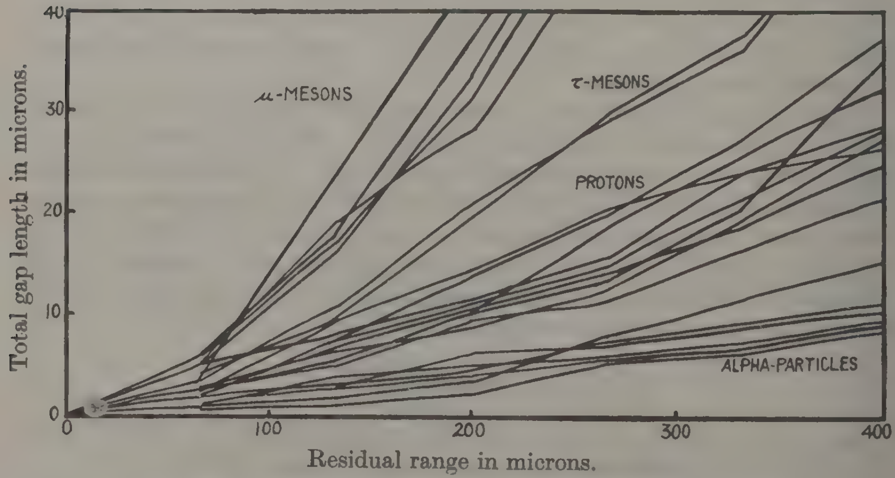
* Communicated by Sir George Thomson, F.R.S.

Fig. 1.



Total gap length against residual range for various particles.

Fig. 2.



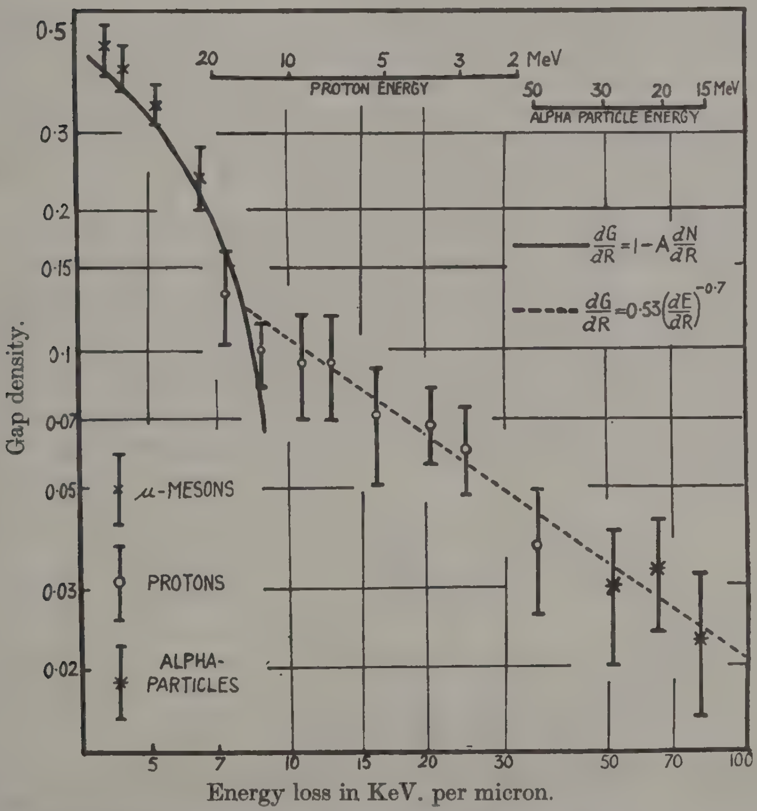
Total gap length against residual range for various particles.

where dE/dR is the energy loss in KeV./μ and dG/dR the gap density. The exact form of this expression depends on the type of emulsion and development, and so each batch of plates must be calibrated.

At lower grain densities, where the clogging is small, the gap density is given by

$$\frac{dG}{dR} = 1 - A\frac{dN}{dR},$$

Fig. 3.



Relation between gap density and energy loss with standard deviations of measurements along 130μ samples of track.

where dN/dR is the grain density in grains per micron empirically related to dE/dR, and $A \sim 0{\cdot}6\mu$ is the mean grain diameter. This curve is included in fig. 3 and agrees with the experimental points for high gap densities. It should be noted that the gap density is a pure number, and so, unlike the grain density, is independent of the size of the eye-piece scale divisions.

The procedure to be followed in order to identify and measure the energy of those particles from cosmic ray stars in electron sensitive emulsions producing clogged tracks depends on whether the particle under investigation ends in the emulsion or not.

If the track ends in the emulsion, a gap measurement is made and the particle identified by comparison with figs. 1 and 2. Visual examination of the track for scattering and δ-rays (Sörensen 1949) confirms the identification. Use of range–energy curves gives the energy of emission of the particles from the star.

If the track does not end in the emulsion, the gap density identifies the particle, provided the track is sufficiently long, except perhaps for some uncertainty between a slow meson and a fast proton, and between a slow proton and a fast alpha-particle. Inspection of scattering in the first case and δ-rays in the second, resolves this difficulty. Using fig. 3, the energy loss can be found from the gap density and hence the energy is known from the energy–energy loss curves.

The above analysis can only be carried out if the track makes a small angle with the plane of the emulsion. The correction factor which must be applied to the gap densities of dipping tracks may be derived from a simple geometrical model. It may be shown that a gap of length g appears of length $(g+\mathrm{A})\cos\theta'-\mathrm{A}$ when viewed from a direction making an angle θ' with the normal to the track. Assuming that the distribution of gap lengths may be replaced by a mean gap length g_m and that the lengths of blocks of grains remains constant during shrinkage, it may further be shown that

$$\frac{d\mathrm{G}}{d\mathrm{R}}=\frac{\cos\theta}{\cos\theta'}\left(1+\frac{\mathrm{A}(1-\cos\theta')}{g'_m}\right)\left(\frac{d\mathrm{G}}{d\mathrm{R}}\right)'+1-\frac{\cos\theta}{\cos\theta'},$$

where $\tan\theta/\tan\theta'=$shrinkage factor$\simeq 2{\cdot}7$,

$$\left(\frac{d\mathrm{G}}{d\mathrm{R}}\right)'=\text{apparent gap density},$$

$$g'_m=\text{mean apparent gap length.}$$

For $\theta<10°$ this correction is small, while for $\theta>20°$, gap measurement is unreliable. The correction factor is large for tracks of low gap density, since the mean gap length decreases with gap density.

This method, although more laborious, is better than simply counting the number of gaps because it gives better discrimination between particles and also it is less sensitive to the angle of dip of the track.

References.

Bradner, H., Smith, F. M., Barkas, W. H., and Bishop, A. S., 1950, *Phys. Rev.*, 77, 462.

Harding, J. B., 1950, *Phil. Mag.*, **41**, 405.

Sörensen, S. O. C., 1949, *Phil. Mag.*, **40**, 947.

A Note on the Use of Resistance Thermometers for Measurement of Rapidly Changing Temperatures.

By C. B. Daish, D. H. Fender and A. J. Woodall.
Military College of Science, Physics Branch *

* Communicated by the Authors.

[Received June 2, 1950.]

In a recent issue of the *Philosophical Magazine* (1950, **41,** 468–477), Leah, Rounthwaite and Bradley describe the use of resistance thermometers in studying the high temperatures produced in explosions. These investigators sought to follow a temperature which rises more than 1000° C. in 20 or 30 milliseconds by using a platinum alloy wire of length 1 inch and diameter 0·0005 inch, and to eliminate radiation errors by a calibration experiment designed to disentangle natural convection from radiation.

In investigations we are making, we have followed the temperature-changes of a long, fine platinum wire suddenly heated and then allowed to cool in an enclosure under conditions in which conduction through the gas is the dominant method of transference of heat. In such conditions the cooling curve of the wire is roughly exponential and experiment substantiates our theory, which assigns to the cooling process a time-constant depending on the thermal conductivity of the gas, on the thermal capacity per unit length of the wire and on the geometry of the system, including the dimensions of the containing vessel. The plain wire used by these investigators would display in such conditions a time-constant of the order of 20 millisec., and the quartz-coated wire would behave appreciably differently. The relatively stout leads would have a time-constant of the order of half a minute, but this will be considerably reduced if natural convection is allowed to come into play.

If this system of wires and leads is exposed to a sudden change in temperature, since convection currents take time to form, the initial heat transfer will be solely by radiation and conduction through the gas and along wires. Conduction to the leads is not negligible. Indeed a calculation shows that in the extreme case when the surface of the wire loses heat by conduction alone, the average temperature of the wire only rises by 93 per cent of the rise at the centre, followed by a subsequent slow drift as the leads warm up. Since the end effects will be very different in the main and the calibration experiments, an error will be thereby introduced, and the effect will moreover be appreciably different for the quartz-coated wire.

Further, the attack on the radiation correction in the explosion experiments relies on a subsidiary experiment on natural convection. As mentioned above, natural convection can have little bearing on the critical first stage of the explosion experiments and the method of obtaining a radiation correction is therefore invalid.

Lastly, but probably of most importance, the true shape of the temperature–time curve must be badly distorted by the lag in response of the wire. The fact that the plain and the quartz-covered wires have different time-constants will ensure that the distortions are not the same in the two cases and this factor alone might conceivably account for the differences observed in the experiments. Certainly the effect of this factor must be very carefully considered before any reliance can be placed on deductions from the curves presented.

It was the conviction that the response of resistance thermometers and thermocouples to sudden changes warranted more attention which prompted us to undertake our investigations some time ago, and we hope shortly to be in a position to make a more adequate communication on the subject.

May 22nd, 1950.

LXV. *Notices of New Books and Periodicals received.*

Relations entre les phénomenes solaires et géophysiques. Méthodes de calcul dans des problèmes de mécanique. Cinétique et mécanisme des réactions d'inflammation et de combustion en phase gazeuse. (Centre National de la Recherche Scientifique : Paris.)

THESE are records of three of the international colloquia which have been organized in recent years with the help of the French Scientific Research Centre.

The first, held at Lyons in 1947, dealt with solar and geophysical phenomena. A main problem in this field is to establish the complex relations between the incident solar radiations and effects observable at the earth's surface. Some evidence is obtainable by balloon and, more recently, by rocket flights, but most of the experimental material presented here is of radio observation on the ionosphere and optical investigation of such phenomena as the aurora and the luminosity of the night sky. On the theoretical side, there is a paper by S. K. Mitra on unsolved problems in the field and articles by H. S. W. Massey and Marcel Nicolet on the photochemistry of the high atmosphere.

The second volume is divided into two parts referring to meetings at Marseilles and Paris in April 1948. The first part contains an exposition by D. N. de G. Allen of Southwell's relaxation method, and an account by T. Vogel of the escalator method of Morris and Head for the determination of normal vibrations. The second is devoted to problems of fluid flow, including in particular articles by J. M. Burgers on turbulent flow, and by L. Malavard on the use of electrical analogues.

The third volume records the colloquium on the initiation and propagation of combustion in gases which was held in Paris at the end of April 1948. It is concerned for the most part with the kinetics and detailed molecular mechanism of the reactions involved. It would give a wrong impression to cite individual articles. Considerable advance has been made in this subject during and since the war, and this colloquium contains much interesting and valuable information.

The benefit of colloquia such as these is greatest for those fortunate enough to take part, but these printed records will be of interest to many not so experienced in such special field as to have contributed thereto. G. W.

The Chemical Elements and their Compounds. By N. V. Sidgwick. Two vols. [Pp. xxxii+1703.] (Oxford: Clarendon Press, 1950.) Price 70*s.*

This work stresses the basic principles underlying the reactions and properties of the chemical elements and compounds. A simple and sensible layout based on the classification of the periodic table, and the emphasis placed on the relations between elements of a group and the groups themselves, provide a clear and satisfying picture of chemistry as a coherent whole. Perhaps the most striking feature of the discourse is the rational treatment accorded to the organic and metal-organic compounds, which are carefully fitted into the pattern instead of being relegated to more highly specialized works. Technical details are reduced to a minimum, with the result that the fundamental structure of chemistry is thrown into relief. Clear exposition combines with an extensive bibliography to make these two volumes a necessity for the chemist's library.

D. H. M.

Non-Linear Vibrations in Mechanical and Electrical Systems. By J. J. Stoker. [Pp. xix+273.] (Interscience Publishers, Inc.) Price 40*s.*

The author is concerned with the free and forced oscillations in two ordinary non-linear differential equations, and although there is no claim to give a survey of the work of the past twenty years, he does take account of much recent work and includes an up-to-date bibliography.

After a discussion of the stability of the first order homogeneous differential equation near a singularity, he considers the forced oscillation in Duffing's Equation, in which a small non-linear element occurs in the restoring force. Sections are included on jump phenomena, that is, discontinuities in the solution as the period of the forcing term is varied and on the appearance of subharmonic oscillations of a permanent type.

The other differential equation is Van der Pol's, in which the damping term is non-linear, so that the damping is negative for small velocities and positive for large velocities. Self-excited oscillations are considered first of all and then the stability of the periodic solutions arising from a small forcing term. A chapter is included on Mathieu's equation and its application to the stability of the solution of Duffing's equation.

The presentation up to this point is elementary and within the reach of non-mathematicians, but to conclude the book there are six appendices proving some results of a more abstruse nature.

K. S.

Les spectres de rayons X et la structure électronique de la matière. By Yvtete Cauchois. (Paris: Centre National de la Recherche Scientifique.)

This short book contains an admirable survey of a subject to which the school of Mlle. Cauchois in Paris has made important contributions. The first part gives a rapid conspectus of the field and discusses especially the influence of chemical binding on the position of the discontinuities in X-ray absorption spectra, while the remaining and greater part of the book deals with the interpretation of the fine structure of these spectra in terms of the permitted electronic energy levels, especially for crystalline absorbers. Emission spectra are also fully treated, though in this field there is a much smaller mass of reliable experimental work.

This subject is of the highest importance as giving substantial information on the behaviour of electrons in molecules and condensed matter, and it is an especial merit of the book to draw attention to the decidedly unsatisfactory state of theoretical work on these questions. Many of the unsolved problems raised are of considerable interest. This book should be read by all interested in the electronic structure of matter.

G. W.

Dynamics of real Fluids. By E. G. Richardson. [Pp. 143.] (London: Edward Arnold, 1950.) Price 21s.

There has for some time been a need for an up-to-date text-book on the physics of real fluids, avoiding both the mathematical complexity of Lamb's classical treatise on ideal fluids, and the engineering details of, say, Dodge and Thomson. One hoped that Dr. Richardson's book would meet this need, and a glance at the chapter headings seemed encouraging. A short introduction on classical hydrodynamics is followed by a long chapter on "Fluids of Small Viscosity" which deals largely with boundary layer theory and turbulent wakes. There follow chapters on compressible fluids, fluids with free surfaces, fluids with anomalous viscosities, and problems involving thermal effects.

As a text-book for students the book seems to the present reviewer to have serious faults. As a summary of the present state of knowledge on the topics treated, the book is probably unique: but to the non-specialist it is very difficult reading. The author's very familiarity with his subject-matter is probably a contributory cause—his researches figure in almost every chapter. The treatment attempted is largely non-mathematical, but equations are sufficiently numerous to necessitate an attempt to understand them, and insufficiently detailed to assist the understanding. The exposition seems too rather obscure in places, and the number of mistakes and misprints is such that more than twenty were noticed at a single reading.

In short the book is excellent in conception, but the execution leaves much to be desired.

N. T.

An Introduction to Heat Transfer. By M. Fishenden and O. A. Saunders. [Pp. 205.] (Oxford: Clarendon Press, 1950.) Price 15*s*.

This book is intended primarily for engineering students but others will also find it a useful summary of much information of practical value. It attempts to explain the fundamental physical principles of the three mechanisms of heat transfer, conduction, convection and radiation, and at the same time to provide quantitative formulæ and numerical data in a form suitable for practical applications. In most cases the basic formulæ are quoted without proof, but are followed by an adequate comparison with available experimental data. Each chapter contains a number of worked examples to illustrate the use of the formulæ.

The first third of the book deals with radiation and conduction in two chapters of about equal length, whilst the remainder is devoted to problems of natural and forced convection. The chapter on thermal conduction contains an account of Schmidt's graphical method of solving problems of unsteady heat flow, and emphasizes the importance of numerical and analogue methods generally. The treatment of thermal convection is based on formulæ which satisfy the principle of dynamical similarity, and is preceded by an elementary discussion of the application of dimensional analysis to convection problems.

The present reviewer feels that it is a valid criticism that this book attempts to cover too large a field in a limited space. In consequence, the treatment of some important aspects of the heat transfer problem is rather brief; for example, forced convection to a gas flowing at high speed is covered in a single paragraph.

E. P. Hicks.

ERRATA.

In the article by K. C. Westfold, of the June issue, there should be added to the References:

Smerd, S. F. and Westfold, K. C., 1949, *Phil. Mag.*, **40,** 831–48.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LXVI. *On the Inhomogeneous Deformation of a Plastic Lamina in a Compression Test.*

By R. Hill, M.A.Ph.D.,
Metal Flow Research Laboratory, Sheffield*.

[Received April 14, 1950.]

Summary.

A theory is formulated for the calculation of stresses and displacements in a thin sheet of arbitrary shape which is being plastically compressed between parallel plates. Account is taken of inhomogeneous distortion caused by friction or edge constraints. By a transformation the solution is made to depend on that for a certain elastic problem of plane strain. Experimental data in support of the theory is provided by compression tests on long rectangular strips with both fluid and solid lubricants. A new method is suggested for estimating coefficients of friction relevant for cold-working processes.

§1. Introduction.

The paper is concerned with the behaviour of a metal lamina compressed plastically between parallel rigid platens†. The term lamina is to be understood as referring to a uniform plane sheet of arbitrary shape, having a thickness small compared with other dimensions. This restriction is made for the sake of simplicity, since the variations of stress and strain through the thickness can then be disregarded provided that the friction on the platens is not too great.

So long as the platens overlap the specimen all parts of the latter are compressed equally and, if there were no friction, the deformation would be uniform and the lateral expansion equal in all directions. The slightest trace of friction, however, causes pronounced inhomogeneous deformation if some dimension of the lamina is very large compared with the thickness. For example, in a long rectangular lamina the compressive stress in the plane needed to overcome the frictional resistance to longitudinal expansion is of order $2\mu Yb/h$, where h is the thickness, $2b$ the length, Y the compressive yield stress, and μ the coefficient of friction. Thus, even when the lubricant is so good that μ is 0·05, the compressive stress in the plane is of order $Y/2$ at the centre of a specimen whose length is ten times the thickness. In a plastically deforming metal this implies that local expansion in the length direction is almost or totally inhibited (the third principal stress being small). In general, a lamina changes its shape appreciably during compression unless it is initially circular (the friction being isotropic). Moreover, the distribution of pressure over the platens is non-uniform, so that the test cannot give a direct measure of the uniaxial stress–strain curve. It is necessary to know the relation between the applied mean pressure $\bar{p}$ and the current yield stress Y in compression. An estimate of $\bar{p}$ for a circular disk, due to Siebel (1923), is

---

* Communicated by the Author.

† Platen is the technical term for a compression plate in a press.

$$\bar{p}=\left(1+\frac{2\mu a}{3h}\right)\mathrm{Y},$$

where a and h are respectively the current radius and thickness. This formula should be a good approximation for a lamina when $\mu a/h$ is not too large, say less than $\frac{1}{2}$. No similar formula has been suggested for a lamina of arbitrary shape, and it is the purpose of the present paper to show how one may be obtained. The change of shape of a non-circular lamina can be regarded as a sensitive indication of the presence of friction. It is certainly a more delicate indication than the barrelling of a solid cylinder, which is not known invariably to accompany friction. Furthermore, by the help of the theory to be described, the coefficient of friction can be deduced from measurements of the dimensional changes during compression.

When, on the other hand, the lamina overlaps the platens, the material outside the platens is not compressed, although a part of it will usually be deformed plastically by the stresses in its plane. A constraint is thereby exerted on the material between the platens, tending to produce an inhomogeneous distortion which may overshadow that caused by friction. In a test suggested by Orowan and described by Ford (1948), advantage is taken of this constraint to secure compression under approximate conditions of plane strain. A rectangular strip is indented by platens completely spanning its width. The material outside the platens is almost entirely rigid and, except near the edges, prevents all strain in the width direction when the compressed area is sufficiently long and narrow. The falling-off in pressure towards the ends of the platens is investigated here. The situation resembles that in the roll-gap in strip-rolling, and the results of the investigation should furnish a method for estimating the state of stress near the edges of the strip.

§2. FRICTIONLESS PLATENS WITH EDGE CONSTRAINT.

We suppose, to begin with, that the platens are perfectly smooth, the non-homogeneous distortion being due solely to constraints applied to the edge of the material between the platens. Although the constraints would in general produce a variation in stress and strain through the thickness, this is confined to the part of the lamina lying within a distance of order h from the edges of the platens*. It can therefore be neglected since we are assuming the width to be large compared with the thickness.

---

* See, for instance, the solution by Hill and Lee (1946) to the problem of the plane indentation of a block by perfectly rough dies.

Let Cartesian axes of reference (x, y, z) be taken, with z normal to the lamina. The components of the mean stress (averaged through the thickness) satisfy the equations of equilibrium :

$$\frac{\partial \sigma_x}{\partial x}+\frac{\partial \tau_{xy}}{\partial y}=0, \qquad \frac{\partial \sigma_y}{\partial y}+\frac{\partial \tau_{xy}}{\partial x}=0. \quad \ldots \ldots \quad (1)$$

Let S denote the speed of approach of the plates at any moment ; the corresponding rate of compression is S/h. The Lévy-Mises plastic stress–strain relations are then

$$\left.\begin{aligned} -\frac{h}{S}\frac{\partial U}{\partial x} &= \frac{\sigma_x-\frac{1}{2}(\sigma_y+\sigma_z)}{\sigma_z-\frac{1}{2}(\sigma_x+\sigma_y)}, \\ -\frac{h}{S}\frac{\partial V}{\partial y} &= \frac{\sigma_y-\frac{1}{2}(\sigma_z+\sigma_x)}{\sigma_z-\frac{1}{2}(\sigma_x+\sigma_y)}, \\ -\frac{h}{S}\left(\frac{\partial U}{\partial y}+\frac{\partial V}{\partial x}\right) &= \frac{3\tau_{xy}}{\sigma_z-\frac{1}{2}(\sigma_x+\sigma_y)}, \end{aligned}\right\} \quad \ldots \ldots \quad (2)$$

where (U, V) are the (x, y) components of velocity averaged through the thickness. We have, finally, the criterion of plastic yielding :

$$\sigma_z=f(\sigma_1, \sigma_2), \quad \ldots \ldots \ldots \ldots \quad (3)$$

where σ_1 and σ_2 are the principal stresses in the plane of the lamina.

Since the lateral strain is non-homogeneous, different elements receive different amounts of work-hardening. However, when the variation in the lateral component of strain is not too great (as we shall assume), all elements are in virtually the same state of work-hardening at any moment. Let Y be the current value of the uniaxial yield stress corresponding to this state. We now take advantage of the fact that the stress in any element will normally not be far from uniaxial compression in the z direction. Then, the hydrostatic component of stress being without influence on yielding, the criterion (3) can be approximated by the linear expression

$$\sigma_z-\tfrac{1}{2}(\sigma_x+\sigma_y)=-mY, \qquad (|\sigma_1-\sigma_2|\leqslant\tfrac{2}{3}mY), \quad \ldots \quad (4)$$

where m is a constant which is to be chosen to give the best overall approximation, and we have used the fact that $\sigma_x+\sigma_y \equiv \sigma_1+\sigma_2$. When the values of the principal stresses are plotted along rectangular axes, the equation (4) represents one face of a regular hexagonal prism whose axis is in the direction (1, 1, 1). When $m=1$, the prism encloses the circular cylinder representing von Mises' criterion and touches it along the generators corresponding to uniaxial tension or compression. By suitably choosing m, equation (4) can therefore be made to approximate von Mises', or indeed any, criterion for states of stress which do not depart far from uniaxial compression*. If m is assigned the value $\sqrt{3}/2$, (4) agrees with von Mises' criterion in a state of pure shear. Hence by choosing m between $\sqrt{3}/2$ and 1, we ensure that (4) is never in error by more than about 8 per cent for any stress in the range $|\sigma_1-\sigma_2|\leqslant 2m\mathrm{Y}/3$. If the shear stress in the plane exceeds $m\mathrm{Y}/3$, the error can only be kept below 8 per cent by replacing (4) by another face of the hexagonal prism. This would not be analytically convenient and we shall therefore take (4) as an approximation for *all* stress states. However, since (4) is introduced mainly to simplify the stress–strain relations (2), the accurate criterion (3) can still be used for an improved estimate of σ_z when approximate values of σ_x and σ_y have been calculated.

---

* Similarly, the maximum shear-stress criterion (represented by a hexagonal prism inscribed in the Mises' cylinder) can sometimes be used as a convenient linearizing approximation to any actual criterion for states of stress near a pure shear.

On inserting (4) and (2), we find that

$$\left.\begin{aligned} \frac{h}{\mathrm{S}}\frac{\partial \mathrm{U}}{\partial x}&=\tfrac{1}{2}+\frac{3(\sigma_x-\sigma_y)}{4m\mathrm{Y}},\\ \frac{h}{\mathrm{S}}\frac{\partial \mathrm{V}}{\partial y}&=\tfrac{1}{2}-\frac{3(\sigma_x-\sigma_y)}{4m\mathrm{Y}},\\ \frac{h}{\mathrm{S}}\left(\frac{\partial \mathrm{U}}{\partial y}+\frac{\partial \mathrm{V}}{\partial x}\right)&=\frac{3\tau_{xy}}{m\mathrm{Y}}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (5)$$

We now substitute

$$u=\frac{m\mathrm{Y}}{\mathrm{E}}\left(\frac{h\mathrm{U}}{\mathrm{S}}-\tfrac{1}{2}x\right), \qquad v=\frac{m\mathrm{Y}}{\mathrm{E}}\left(\frac{h\mathrm{V}}{\mathrm{S}}-\tfrac{1}{2}y\right), \quad . \quad . \quad . \quad (6)$$

where E denotes Young's modulus, and the new variables (u, v) have the dimensions of length. We obtain

$$\mathrm{E}\frac{\partial u}{\partial x}=\tfrac{3}{4}(\sigma_x-\sigma_y)=-\mathrm{E}\frac{\partial v}{\partial y}, \qquad \mathrm{E}\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)=3\tau_{xy}. \quad . \quad . \quad (7)$$

These will be recognized as the *elastic* equations for a material which has a Young's modulus E and Poisson's ratio $\frac{1}{2}$, and which is deformed in *plane strain*. The distribution of stress and strain in the plastic lamina can therefore be derived from that in a geometrically similar lamina of incompressible material deformed elastically in plane strain. The (x, y) stresses are indentical in both, while according to (6), the elastic displacement is proportional to the plastic velocity representing the non-homogeneous part of the strain (*i. e.* over and above uniform compression and expansion). The z component of stress needed to maintain plane elastic strain in an incompressible material is $\frac{1}{2}(\sigma_x+\sigma_y)$; thus, from (4), the pressure exerted by the platens on the plastic lamina is equal to $m\mathrm{Y}$ plus the pressure on the elastic lamina.

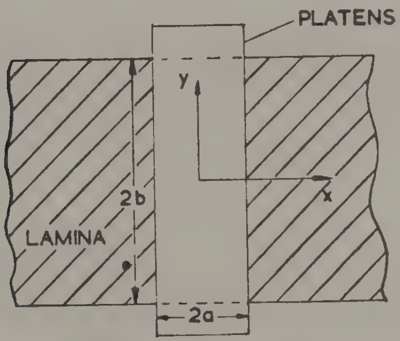
Consider, for example, a long rectangular lamina which is being compressed between smooth platens completely spanning its width (fig. 1). Let the width of the lamina be $2b$ and the width of the platens $2a$. Let the x and y axes be taken parallel to the sides of the lamina, with the origin at the geometric centre of the area under compression. If the material lying outside the platens is assumed to be rigid, the conditions along the boundary of the rectangular compressed area are

$$\mathrm{U}=\text{constant}=\pm\lambda\mathrm{S}a/h,\ \text{say}\ ; \qquad \mathrm{V}=0\ ; \qquad \text{on } x=\pm a\ ;$$

together with zero applied stress on the surfaces $y = \pm b$. The unknown constant λ must lie between $\frac{1}{2}$ (corresponding to free expansion) and 1 (corresponding to complete prevention of lateral strain). The precise value is determined by the condition that the resultant force over a lateral section, acting in the x direction, is zero (the ends of the lamina being unconstrained). The boundary-conditions in the associated elastic problem for a rectangular lamina (with sides $2a$ and $2b$) are

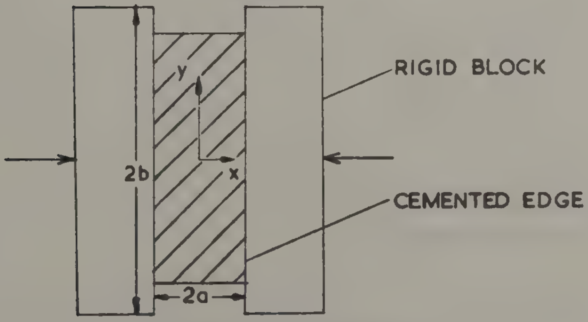
$$u = \text{constant} = \pm(\lambda - \tfrac{1}{2})m\mathrm{Y}a/\mathrm{E}, \qquad v = -m\mathrm{Y}y/2\mathrm{E}, \qquad \text{on } x = \pm a\,;$$

Fig. 1.



Lamina compressed between platens which overlap across the width only.

Fig. 2.



Associated elastic problem.

together with zero applied stress along $y = \pm b$ and zero resultant force on the surfaces $x = \pm a$. These conditions are satisfied by superposing a uniform tensile stress $2m\mathrm{Y}/3$ in the x direction on the stress distribution due to a compressive load $4m\mathrm{Y}b/3$ applied through inextensible *rigid* bars *cemented* along $x = \pm a$ (fig. 2). According to (7) the tensile stress produces displacements

$$\pm m\mathrm{Y}a/2\mathrm{E}, \qquad -m\mathrm{Y}y/2\mathrm{E}, \qquad \text{on } x = \pm a.$$

The compressive load balances this tension (so satisfying the condition for zero resultant force), and produces along $x = \pm a$ a zero displacement in the y direction (since the bars are cemented and inextensible) and a certain uniform displacement in the x direction (since the bars are rigid). The magnitude of the latter has to equal $(1-\lambda)\, m\mathrm{Y}a/\mathrm{E}$, and this determines the value of λ.

The problem of compressing a rectangular elastic block between massive cemented bars does not appear to have been solved*. A photo-elastic investigation is projected, though the achievement of plane strain presents difficulties. We may, however, safely make certain generalizations about the distribution of stress, without a detailed solution. In the first place, by Saint-Venant's principle, the stress must be virtually uniform except within a distance of order $2a$ from the free surfaces. Secondly, the pressure transmitted through the bars must fall off at the edges, since the material there has greater freedom to expand and can therefore be compressed by a given amount under a smaller pressure. Thirdly, since the bars are cemented, they exert a shearing stress and this, by Saint-Venant's principle, must be concentrated mainly within a distance of order $2a$ from the free surfaces. Translating these conclusions in terms of the plastic lamina, we can expect a state of plane strain except in a region near the edge, where there is both a shearing stress and a *tension* parallel to the edge. The pressure applied through the platens therefore diminishes towards the edge, the fall-off being confined to a region whose width is roughly equal to that of the platens.

§3. ROUGH PLATENS WITHOUT EDGE CONSTRAINT.

We consider, now, a lamina free from edge constraints and compressed by rough platens which overlap it all round its circumference. The coefficient of friction μ is assumed so small that terms containing second or higher powers of μ can be neglected. For example, σ_z is a principal stress to first-order, and so the yield criterion is still (3) or (4). The theory is only a valid account of a first-order departure from pure uniform compression. This is the reason for excluding edge constraints.

It is necessary to determine which element of the lamina remains momentarily at rest at each stage of a slow compression. In a fast compression, where inertial forces are large compared with frictional forces, it is clearly the element momentarily coinciding with the centroid, since there are no other lateral external forces. The situation is different, however, when the compression is so slow that inertial forces are negligible compared with frictional forces. For a first-order calculation of the distribution of the frictional shear stress over the faces of the lamina, the deformation may be taken to be uniform expansion. We must also allow for the possibility of a small rotation of the lamina in its plane. Then, to first order, the relative velocity of an element distant r from the momentarily fixed point is compounded of a radial velocity proportional to r (due to the uniform expansion) and a velocity which is perpendicular to this and also proportional to r (due to the rigid-body rotation). The resultant velocity is therefore proportional to r and inclined at a fixed angle θ to the radius. Hence the frictional stress, opposing the relative motion, is directed at the same constant angle θ to the radius. To first order, the magnitude of the frictional stresses is μY, and they therefore exert no resultant torque on the lamina if

---

* An approximate method, yielding limited information and applied only to a square block, has been described by H. J. Greenberg and R. Truell (1948).

$$\mu Y \sin\theta \iint r\,dx\,dy = 0,$$

taken over the lamina. Since the integral is essentially positive, it follows that $\theta=0$, or that there is no rotation of the lamina as a whole. The friction is therefore directed toward the point at rest, and the condition for zero resultant frictional force is

$$\iint \mathbf{t}\,dx\,dy=0, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (8)$$

where $\mathbf{t}$ is the unit vector in the radial direction. This is the equation determining the position of the element momentarily at rest. If the lamina has two axes of symmetry, the centroidal element always remains at rest. For a lamina of arbitrary shape, however, the point defined by (8) is not the centroid.

Let (x, y) axes be taken with the origin coinciding with the point at rest. The frictional stress contributes to the equations (1) additional terms which are equivalent to a body stress with potential $2\mu Yr/h$. Such a body stress can be balanced by a variable hydrostatic tension whose value is equal to the potential. The remaining part of the actual stress then satisfies the equations (1) of equilibrium. Furthermore, the remaining part of the actual stress still satisfies the plastic stress–strain relations and yield criterion, since these are unaffected by the addition of a hydrostatic stress. The corresponding external stress on the edge of the lamina must be equal to a normal pressure of amount $2\mu Yr/h$ at each point in order to produce, when combined with the hydrostatic tension, zero resultant stress on the edge.

Thus, to obtain the distribution of stress in a lamina compressed by rough platens, we combine a varying hydrostatic tension $2\mu Yr/h$ with the distribution of stress in a similar lamina compressed along its faces by smooth platens and along its edge by a varying normal pressure $2\mu Yr/h$. The deformation is identical in the two laminæ. Since the edge pressure is proportional to the distance from the point at rest, it offers the least resistance to lateral expansion of the parts of the edge nearest to this point. In other words, as one would naturally expect, friction on the platens causes the lamina to expand most in the direction of its least dimension.

The elastic analogy described in §2 is, of course, still applicable. Let us, for example, apply it to determine the stress in a circular lamina of radius a. The corresponding edge pressure produces a uniform hydrostatic pressure $2\mu Y a/h$ in the plane of the elastic lamina and hence also in the plastic lamina between smooth platens. Combining this with the varying hydrostatic tension $2\mu Y r/h$, we obtain equal radial and circumferential compressive stresses $2\mu Y\ (a-r)/h$ in the plane of the lamina between rough platens. From (4), taking $m=1$, the pressure exerted by the platens is

$$Y+2\mu Y(a-r)/h$$

and the mean pressure is $Y\ (1+2\mu a/3h)$, in agreement with Siebel's formula.

§4. EXPERIMENTAL DETERMINATION OF THE COEFFICIENT OF FRICTION.

The preceding theory provides a means of estimating the coefficient of friction in a compression test from observations of the change in shape of a rectangular lamina.

The coefficient of friction, so determined, corresponds to the conditions in a compression test, viz: where one of the sliding bodies is distorted elastically in bulk and the other has yielded plastically. Such coefficients of friction are relevant for all processes for cold-working metals .

The method is obviously more sensitive when the change of shape is large, but then the first-order approximation of the theory is less accurate. However, when the lamina is long and narrow, a more exact calculation can be made. If the length $2b$ is greater than about 10 times the width $2a$, and if a/h is of order unity, it is probably legitimate to neglect the variation in stress across the width in comparison with the variations along the length. The only stress components are then the longitudinal compressive stress q, the pressure p, and the friction μp. To the same order of approximation we may disregard the change of *shape* when setting up the equation of equilibrium. This then reduces to

$$h\frac{dq}{dy}=2\mu p, \quad \ldots \ldots \ldots \quad (9)$$

where y is measured along the lamina from its shorter side $2a$. For any yield criterion near that of von Mises, it is sufficiently accurate to use equation (4):

$$p-\tfrac{1}{2}q=mY. \quad \ldots \ldots \ldots \quad (10)$$

Eliminating q from (9), and integrating, we obtain

$$p=mY\exp(\mu y/h), \quad \ldots \ldots \ldots \quad (11)$$

after using the boundary-condition $p=mY$ on $y=0$.

The longitudinal strain-increment vanishes at a section distant y' from the shorter side, such that $q=\tfrac{1}{2}p$ or

$$\mu y'/h=\ln(4/3)\simeq 0{\cdot}288. \quad \ldots \ldots \ldots \quad (12)$$

Hence, if $\mu b/h < 0{\cdot}288$, all parts of the specimen continue to elongate. It follows from equations (2), (10) and (11), that the increase in b during a further small *decrease* dh in thickness is

$$db = \frac{dh}{h}\int_0^b \left(2 - \tfrac{3}{2}\, e^{\mu x/h}\right) dx = dh\left[\frac{2b}{h} - \frac{3}{2\mu}\left(e^{\mu b/h} - 1\right)\right].$$

This may be written to the present order of approximation as

$$\frac{db}{b} \simeq \left(1 - \frac{3\mu b}{2h}\right)\frac{dh}{2h}.$$

Integration gives

$$\frac{b}{b_0} = \left(\frac{1 - \dfrac{\mu b}{2h}}{1 - \dfrac{\mu b_0}{2h_0}}\right)\sqrt{\left(\frac{h_0}{h}\right)}; \quad \frac{\mu b}{h} \leqslant 0{\cdot}288; \quad . \quad . \quad . \quad . \quad (13)$$

where b_0 and h_0 are initial values. The first factor in this equation becomes increasingly less than unity as the compression proceeds, and represents the effect of friction in hindering changes of length. The corresponding formula for the mean pressure is, with $m \sim 1$,

$$\bar{p} \simeq \left(1 + \frac{\mu b}{2h}\right)\mathrm{Y}; \quad \frac{\mu b}{h} \leqslant 0{\cdot}288. \quad . \quad . \quad . \quad . \quad . \quad (14)$$

On the other hand, when $\mu b/h > 0{\cdot}288$, a central region of the lamina, of length $2(b - y')$, momentarily undergoes no extension. Since there is no relative longitudinal slip in this region, p has the uniform value $4m\mathrm{Y}/3$ and q is equal to $2m\mathrm{Y}/3$ (m should be taken nearer to $\sqrt{3}/2$ as y'/b becomes progressively less). The mean pressure is easily found to be

$$\bar{p} \simeq \left(\frac{4}{3} - 0{\cdot}051\,\frac{h}{\mu b}\right) m\mathrm{Y}; \quad \frac{\mu b}{h} \geqslant 0{\cdot}288. \quad . \quad . \quad . \quad . \quad (15)$$

The increase in length during a small decrease dh in thickness is given by

$$\frac{db}{y'} = \left(4 - \frac{h}{\mu y'}\right)\frac{dh}{2h},$$

or

$$\frac{db}{dh} \simeq \frac{0{\cdot}076}{\mu}; \quad \frac{\mu b}{h} \geqslant 0{\cdot}288. \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

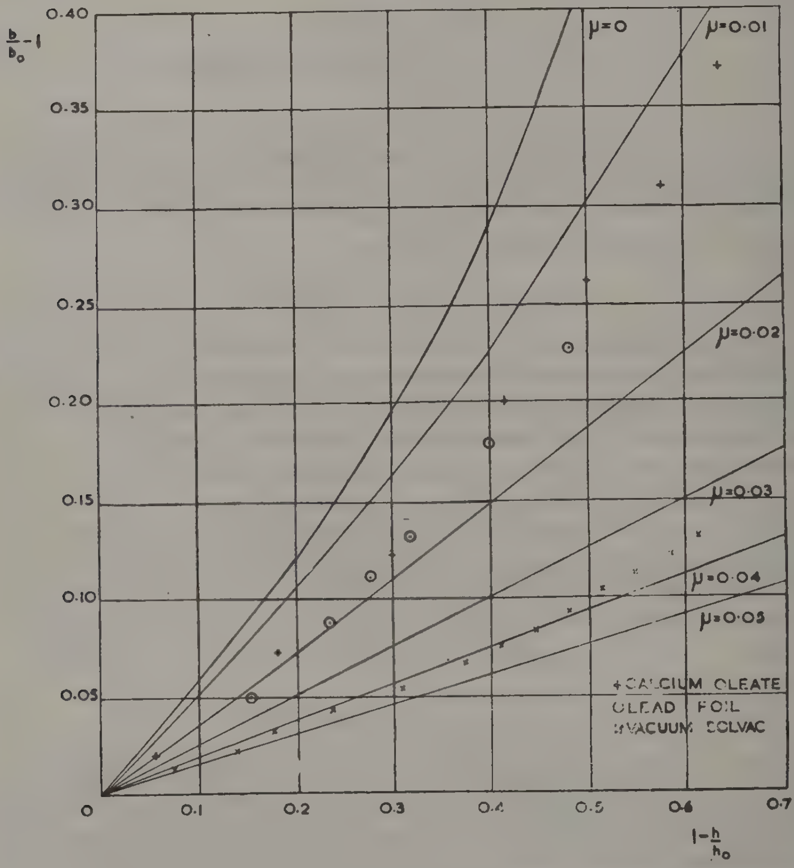
In the present experiments the nominal specimen dimensions were $1{\cdot}0 \times 0{\cdot}1 \times 0{\cdot}05$ in., so that $b_0/h_0 = 10$. Fig. 3 shows the corresponding theoretical relation between the fractional change in length $(b/b_0 - 1)$ and the fractional reduction in height $(1 - h/h_0)$ for various values of μ. At the critical value of b/h, for which the centre of the lamina undergoes no extension, the curve (13) is replaced by a line of slope $0{\cdot}076\, h_0/\mu b_0$, according to (16). When μ is greater than $0{\cdot}0288$ the relation is linear from the start when $b_0/h_0 = 10$.

Also shown in fig. 3 are experimental points for two lubricants: powdered calcium oleate and vacuum solvac. Copper specimens were used, the respective dimensions being

$1{\cdot}000 \times 0{\cdot}100 \times 0{\cdot}050$ in. and $1{\cdot}002 \times 0{\cdot}103 \times 0{\cdot}049$ in.

The copper was rendered isotropic by suitable rolling and heat treatment; this was confirmed by compression tests on circular disks, which remained effectively circular The lubrication was renewed whenever the test was interrupted to take measurements of length and thickness. The experimental points follow the trend of the calculated curves, thus affording indirect evidence that the theory is sound. The deduced coefficients of friction are 0·015 and 0·045, respectively. The latter is comparable with values inferred from measurements of roll-force when rolling *short* lengths of copper strip with vacuum solvac as a flood lubricant*.

Fig. 3.



Comparison of calculated and observed values of the extension of laminæ of dimensions 1·0×0·1×0·05 ins.

---

* Private communication from Mr. R. B. Sims, Head of the Rolling Section, British Iron and Steel Research Association.

The low coefficient of friction obtained with calcium oleate indicates that true boundary-lubrication was not operative, and this was confirmed by the appearance of the copper after compression. The surface was still mainly in the 'orange-peel' condition, in marked contrast to the smoothed and burnished surface of the specimen lubricated with vacuum solvac.

For a direct check of the theory it is necessary to use a lubricant whose frictional properties are known independently.

A *sufficiently thin* foil of some softer metal acts as an efficient lubricating agent, the frictional stress being equal to the limiting yield stress s in shear of the foil at large strains. The foil must not be too thick, or it will extrude in the width direction before (or while) the specimen itself is compressed, and so exert a *negative* frictional drag in that direction. A rough criterion, ensuring non-extrusion, is that Y/s should be less than w/t, where t is the foil thickness and w is the specimen width. This follows from the condition that the hydrostatic component of the stress at the middle of the foil, necessary to cause extrusion in the width direction, should be greater than Y. For lead foil on copper, Y/s is less than about 25/0·5 (units of tn./in^2.) at all strains, and so t/w should be less than 1/50. In the present experiment t was not more than 0·002 in., while w was initially 0·1 in. The foil was removed and replaced by a new piece each time measurements were taken. The results are shown in fig. 3 and correspond to an apparent coefficient of friction of about 0·017 (the changes needed in the theory when the frictional stress is constant are insignificant here). By comparison s/Y is about 0·4/20 or 0·02 for large strains. The agreement is excellent in view of the various uncertainties. Similar confirmation was obtained with pure lead foil on aluminium where the apparent coefficient of friction was found to be 0·042 ($s/Y \sim 0{\cdot}4/10 = 0{\cdot}04$), and with indium foil on a lead alloy where the coefficient of friction was 0·033 ($s/Y \sim 0{\cdot}1/3 = 0{\cdot}033$). The limiting values of Y at large strains (greater than 20 per cent compression) were obtained from measurements of the load and thickness during the test (the current area being calculated as initial volume divided by current thickness). The limiting shear stress of the indium foil was inferred from hardness tests on the bulk metal. An ⅛ in. diameter ball under a 4 lb. load of 20 seconds duration gave an impression diameter of 0·066 in. (average of three tests). This corresponds to a Meyer hardness of 0·52 tn./in^2. and hence to a mean uniaxial yield stress of about 0·17 tn./in^2. over the range of strains produced around a ball indentation.

For indium foil on copper and aluminium, or lead foil on steel, extrusion must have occurred since the change in length was much smaller than would be expected if the foil had behaved as a true lubricant; extrusion would, indeed, be expected by the criterion mentioned previously. In agreement with this hypothesis, a solid cylinder of copper showed barrelling when lubricated with lead, but a concavity when lubricated with indium.

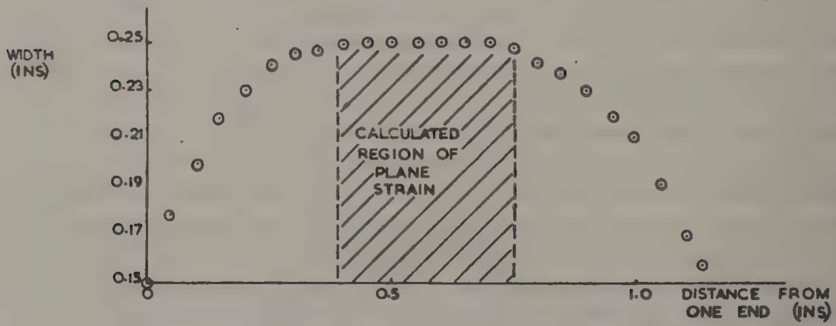
A further check on the theory is provided by a comparison of the observed and calculated values of the length of that part of the specimen which is deformed *throughout* in plane strain. Theoretically this is, from (12),

$$2(b_0-y_0')=2b_0\left(1-\frac{0\cdot288\,h_0}{\mu b_0}\right),\quad \text{when}\quad \mu b_0/h_0\geqslant 0\cdot288.$$

For the copper lubricated with vacuum solvac, where $b_0=0\cdot501$, $h_0=0\cdot049$, $\mu=0\cdot045$, this unextended central region is calculated to be of length 0·37 in. This is indicated on fig. 4, which shows the measured profile of the lamina after a reduction in thickness of 60 per cent. The agreement is excellent. The final maximum width is 0·248 in., while the width calculated on the assumption of plane strain is 0·26 in. When calcium oleate was the lubricant the profile was completely rounded, again in agreement with theory since $\mu b_0/h_0=0\cdot15<0\cdot288$.

It should be carefully noted that the theory becomes less accurate as the lamina is progressively compressed, since the length/width ratio decreases from the initial value of 10 to about 5 after a 50 per cent reduction. Tests on laminæ with an initial length/width ratio of 4 showed that the theory significantly underestimated the increase in length; this was to be expected from the neglect of the component of the frictional forces restraining the expansion in the width direction. This effect can be recognized in fig. 3, where the upward trend of the experimental results at large reductions is more pronounced than the theory predicts. For this reason the coefficients of friction were deduced from the experimental results below about 40 per cent, where the length/width ratio is greater than 7.

Fig. 4.



Profile after 60 per cent reduction with vacuum solvac.

Acknowledgments.

I am grateful to the Director of the British Iron and Steel Research Association for permission to publish this paper; to Mr. N. Haslam of the Metzllurgy Section at the Metal Flow Research Laboratory, Sheffield, for carrying out the experiments; and to Dr. D. Tabor for helpful discussions.

References.

Ford, H., 1948, *Proc. Inst. Mech. Eng.*, **159,** 115.
Greenberg, H. J., and Truell, R., 1948, *Quart. App. Maths.*, **6,** 53.
Hill, R., and Lee, E. H., 1946, *Proc. 6th Int. Cong. App. Mech.* (Paris).
Siebel, E., 1923, *Stahl und Eisen*, **43,** 1295.

LXVII. *A Theoretical Investigation of the Effect of Specimen Size in the Measurement of Hardness.*

By R. Hill, M.A., Ph.D.,
Metal Flow Research Laboratory, Sheffield*.

[Received April 25, 1950.]

Summary.

The theory of plane plastic strain due to Hencky and Geiringer is applied to calculating the minimum dimensions that a specimen should have for a valid hardness test. Critical widths and thicknesses are determined for rectangular specimens indented by flat dies and wedges of any angle. The manner in which the specimen deforms during the indentation is considered in detail.

§1. Introduction.

The definition of indentation hardness adopted in the present paper is the ratio of the applied load to the projected area of the impression. It is equal to the work done per additional unit volume of indentation, or, when the frictional resistance is negligible, to the mean normal pressure between indenter and medium. It is well known that the hardness is not observably dependent on the size of the specimen provided this is greater than a certain minimum (other things being equal). The minimum size depends mainly on the properties of the material under test and on the shape of the indenter. If the specimen is too thin or too narrow, the plastic zone penetrates either to the rear or lateral surfaces; this is accompanied by a change in the apparent hardness (usually a reduction).

The minimum width and thickness of a rectangular specimen are calculated here for indentation by a wedge. At present, this is the only indenter for which a reasonably accurate calculation of the surface distortion and stress distribution is feasible. The problem is treated as two-dimensional, the strain being zero in the direction made by the intersecting faces of the wedge. The material is supposed to be rigid up to the yield point, after which the maximum shear stress retains a constant value in plane strain.

§2. Criterion for the Critical Thickness.

The slip-line field[†] ABCDE in the region of plastically-deforming material around a wedge indenting the plane surface of a semi-infinite mass is shown in fig. 1 (Hill, Lee, and Tupper 1947). The wedge angle is

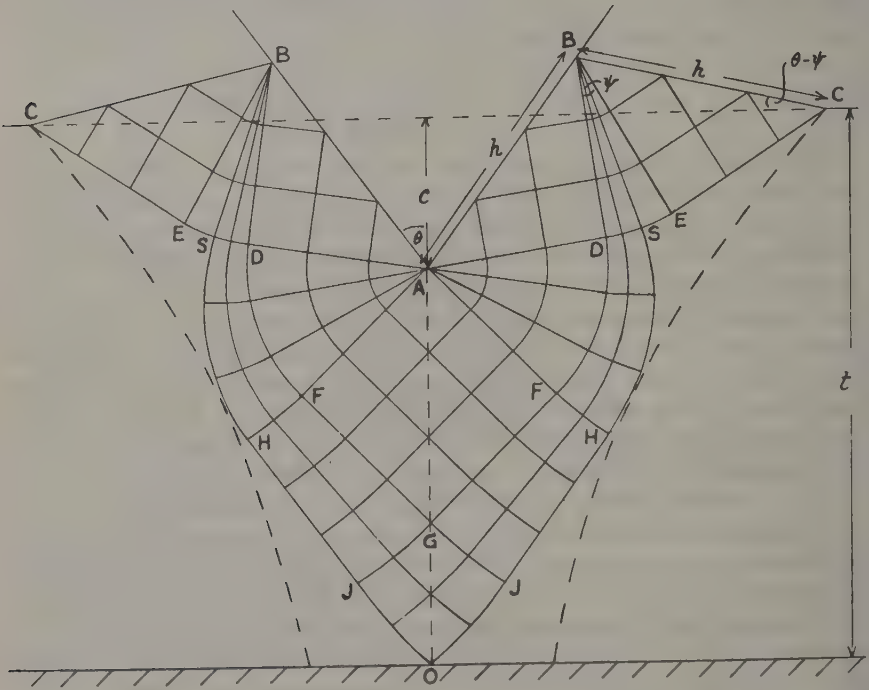
---

[*] Communicated by the Author.

[†] For an account of the theory of plane plastic strain, see the writer's *The Mathematical Theory of Plasticity*, Chap. VI. (Oxford: Clarendon Press, 1950).

denoted by 2θ, the depth of penetration by c, and the length of the line of contact by h. The displaced surface is straight, and, when the wedge is smooth (as assumed here), the coronet is of length h. The inclination of the displaced surface to the horizontal is denoted by $\theta-\psi$; such a raised coronet is characteristic of a cold-worked metal (Heyer 1937). Fig. 2 shows the relations between $h \cos \theta/c$, ψ and θ. When the specimen is semi-infinite and homogeneous the configuration remains geometrically similar and the hardness is independent of load. This may be realized in practice, for example in a Vickers pyramid hardness test, provided the impression is large compared with the grain-size of the specimen and the radius of curvature of the tip of the indenter (Field Foster 1948).

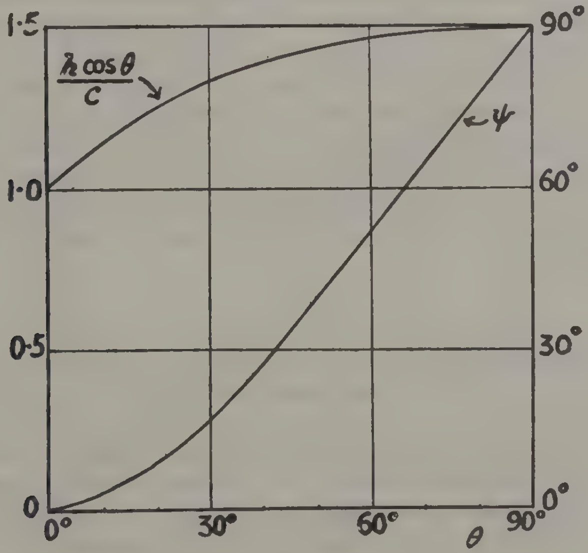
Fig. 1.



ABCDE does not constitute the whole plastic region. Beyond this field lies a further plastic zone where the material is stressed to the yield-point but remains rigid, being prevented from deforming by the constraint of the non-plastic matrix (Hill 1949). The state of stress and strain in the material undergoing deformation is independent of the distribution of stress in the rigid material. Consequently, there is no need to calculate the latter whenever it is evident that the rigid material is capable of supporting the stresses acting around its perimeter. This is necessarily so when a specimen is sufficiently large compared with the localized area of deformation. It follows that conditions in the plastically-deforming zone are independent of the size of the specimen, provided only that it is sufficiently great. In particular, the configuration around a wedge remains geometrically similar (though the shape of the plastic boundary does not).

These conclusions are strictly valid only for a material that remains *rigid* whenever the maximum shear stress is less than k. In an actual solid, where Young's modulus E is finite, the deformation in the plastic region is dependent on the elastic distortion elsewhere. For example, the volume of the coronet is governed by the extent to which the elastic resilience of the whole specimen can accommodate the displaced material. Such departures from the behaviour of a plastic-rigid body are presumably only of order k/E, and the hardness should be independent of specimen size over a certain minimum, within normal experimental error (as is observed).

Fig. 2.



In a specimen whose thickness is small compared with its width, the plastic boundary moves relatively further from the indenter and nearer to the rear surface as indentation proceeds. When the specimen rests on a plane rigid foundation, it is a natural supposition (and one which is suggested by observations of the plastic distortion of the rear surface) that the relative expansion of the plastic zone is mainly concentrated directly below the indenter. At some stage of the indentation the element situated at the midpoint O of the rear surface is stressed to the yield limit. At a little later stage the plastic region contains all the material within a certain area which lies below the indenter and intersects the rear surface over a finite length (the conjectured plastic boundary at this stage is indicated by broken curves in fig. 1). It may be proved (Hill 1949) that deformation is confined to the area ABCDE up to the moment when the plastic region has spread so far that it just includes the two slip-lines OS from O. Thereupon the constraint fails, the plastic core deforms, and the rigid ends are thrust apart, sliding over the foundation. The configuration round a wedge no longer remains geometrically similar and subsequent hardness measurements are spurious.

We now examine the criterion determining the critical depth of penetration. The slip-line field bounded by OS and AS is uniquely determined by the known shape of AS ; the particular position of S on ADEC depends, of course, on the ratio of AO to AB. The field having been calculated, Hencky's theorem determines the normal stresses acting across AO, and hence the total lateral thrust exerted on each half of the specimen by the stresses along OAB. The thrust vanishes when the foundation is smooth and the ends of the specimen are unconstrained. Now the field OSAS could always be formally constructed at any stage of the indentation, but only for one particular depth of penetration would the corresponding thrust across OAB vanish. Before this depth is attained, the plastic region can therefore not have spread sufficiently far to enclose the entire length of the slip-lines OS. Hence, as already stated, the material below the indenter must have remained rigid. The criterion for the critical penetration is therefore that there should be no resultant lateral thrust over OAB corresponding to a formal construction of the field OSAS.

§3. Calculation of the Critical Thickness.

Slip-line fields with a 5° mesh were constructed for smooth wedges of 60° and 120° total angle by the numerical method due to Hill and Lee (1946). The field OSAS consists of a region ADF where the slip-lines are radii through A and circular arcs ; a region DFHS defined by the circular arcs DF and DS of equal radius ; a uniformly stressed region AFG ; a region FGHJ where one family of slip-lines are straight ; and a region GJO defined by the slip-line GJ and the axis of symmetry.

If t is the minimum specimen thickness, it is found that $t/2h$ is equal to 1·04 for $\theta=30°$ and to 2·17 for $\theta=60°$. The corresponding distributions of normal stress across OA are shown in fig. 3 ; it will be observed that this stress is tensile immediately under a narrow wedge but compressive under a wide wedge. The stress in the rear surface is always tensile. When θ tends to zero, so that the indenter becomes a knife-edge, the pressure $2k$ along AB is balanced by a uniform tension πk across OA. $t/2h$ is therefore equal to $(\pi+2)/2\pi$ or 0·82 approximately. At the other extreme, when θ tends to 90°, the slip-line field becomes similar to that for a flat die, for which the critical $t/2h$ is 4·37 (Hill 1947). The relation between the critical ratio $t/2h$ and the semi-angle θ is given in fig. 4*. The critical ratio of load to thickness is also shown; since this has a maximum value when $\theta \sim 60°$, the hardness of a given specimen can be measured over a greater range of loads with a 120° wedge than with any other.

Fig. 3.

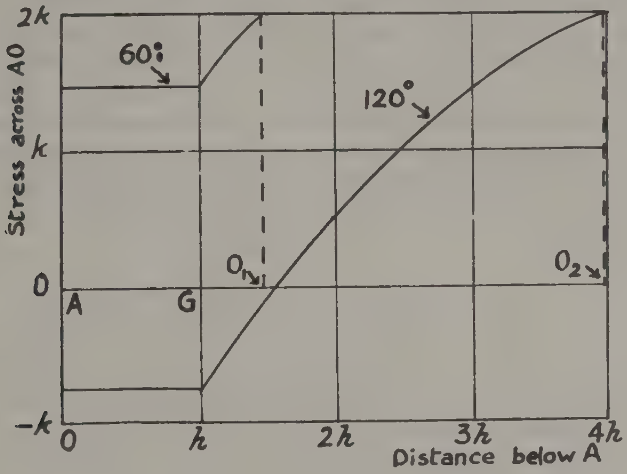
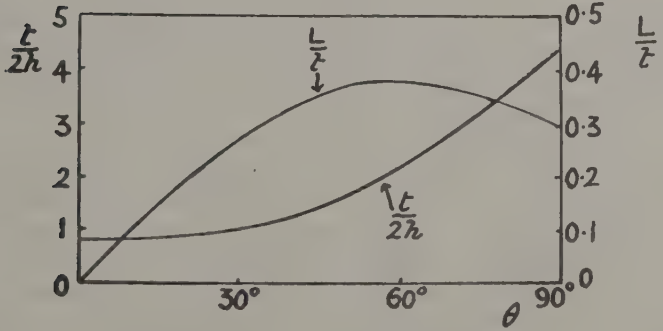


Fig. 4.



The effect of friction on the foundation is to delay the moment when the critical penetration is reached. Calculation shows, however, that the effect is so small as to have little practical importance for coefficients of friction less than 0·1. Lateral compressive forces applied to the sides of the specimen would be more effective (so long as they do not cause general yielding).

---

* For comparison, the critical ratio of specimen thickness to impression diameter is 3·14 when a semi-cylindrical cavity is expanded in the surface (Hill 1948).

The speed U with which the ends of the specimen slide sideways is determined by the following considerations. If the speed V of penetration is given, and U is assigned arbitrarily, the velocity at any point in the plastic region can be calculated from Geiringer's equations. In particular, the velocity in AFG, ABD, and BEC is uniform, so that the displaced surface BC remains straight during the next infinitesimal increment of penetration. (As the penetration continues, the part of the coronet above the element now at C remains straight, but the additional lower part is curved.) Corresponding to a slightly altered position of the surface, it must be possible to find a new plastic state of stress throughout the material which was plastic prior to the increment of penetration. Otherwise plastic material unloads during the incremental penetration and this is inconsistent with the assumption of continuing plastic flow implicit in the application of Geiringer's equations*. This is the condition which determines the true ratio U/V. A calculation of U/V has not been made by the writer.

§4. CALCULATION OF THE CRITICAL WIDTH.

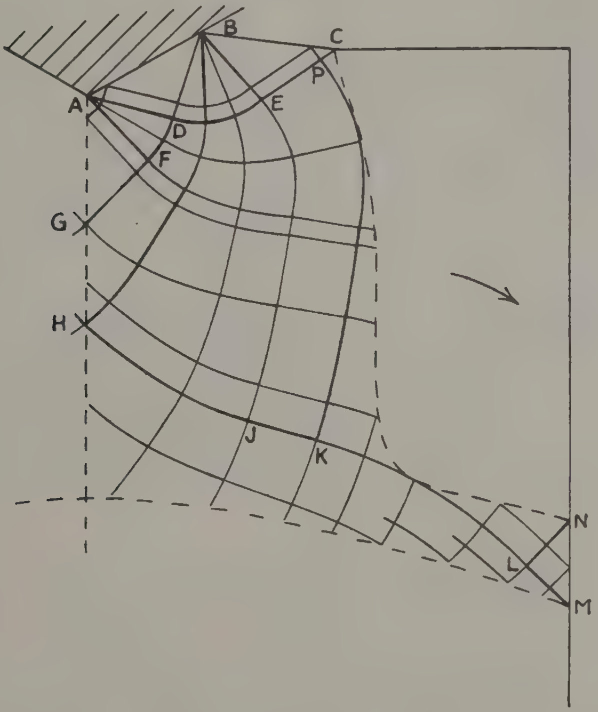
In a specimen whose width is small compared with its thickness, the plastic region spreads out to the sides as indentation continues. At a certain penetration the plastic region (whose conjectured boundary is indicated by broken curves in fig. 5) just contains a complete slip-line HJKLM, extending from the axis of symmetry to the lateral surface (which it must meet at an angle of 45°). Up to this moment the plastic material below the indenter has been rigidly constrained. After this moment the rigid corners of the specimen are assumed to be displaced sideways by sliding over the slip-line KLM; for this to be possible, KLM must be of a suitable shape. Now the development of a plastic region having a boundary cut more than once by a slip-line family is not uniquely determined by the state of stress alone, but proceeds in such a way that certain conditions for the existence of displacements are satisfied (Hill 1950). It is reasonable to suppose, therefore, that the plastic region spreads out to the sides of the specimen in such a manner that the slip-line KLM is of the correct shape to permit a sideways displacement of the rigid corners. Since a rigid-body motion in a plane can only be a rotation about some instantaneous centre (which is at infinity for a pure translation), two rigid bodies can only slide over each other if the line of contact is a circular arc. Thus KL must be circular. We must also include a straight segment LM such that LMN is a region with straight slip-lines meeting the surface at 45°. The stress in LMN is a uniform compression $2k$ parallel to the surface, while the ensuing deformation caused by the rotation of the corner is a non-uniform shear parallel to LM.

---

* These remarks are of general significance for removing an apparent lack of uniqueness in many problems where a plastic-rigid body is assumed.

Since the nett angle turned through in going from M to H along the slip-line is zero, the hydrostatic components of the stress at H and M are equal; their value is $-k$, since the stress in LMN is a compression of amount $2k$. Hence H is the point on the axis at which the hydrostatic stress-component is $-k$ in the unique continuation of the field ABCDE. By Hencky's theorem, H lies on the slip-line through B which is inclined at an angle $\frac{1}{4}\pi-\frac{1}{2}(\theta-\psi)$ to BE. For θ less than about 50°, H coincides with A, and the slip-line HJKLM makes an angle $\frac{1}{4}\pi-\frac{1}{2}(\theta-\psi)$ with AD.

Fig. 5.

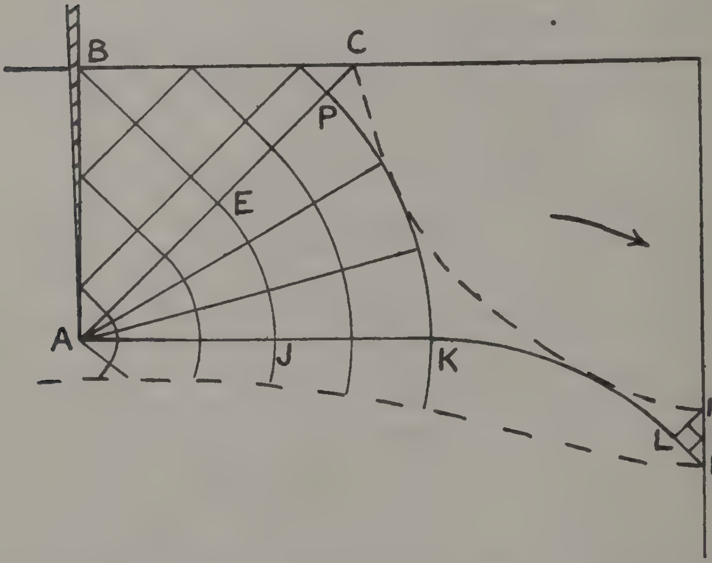


H having been fixed in this way, there are three unknown quantities to be calculated. These are the length b of JK (all slip-lines of this family of EJKP are of equal length), the radius r of KL, and the length a of LM. When these are known, the critical width of the specimen follows at once from simple geometry. To determine the three unknowns there are the three conditions for the equilibrium of the rigid corner, viz.: two conditions for the vanishing of the resultant force components, and one for the vanishing of the resultant couple. The resultant force and couple are calculated from the distribution of stress over the perimeter CPKLN, or, more conveniently, from the distribution over BAHJKLM (the equilibrium of the region covered by the slip-line field is, of course, automatically ensured by the equations of Hencky used in constructing the field).

When the indenter is a knife-edge an explicit analytic solution is possible (fig. 6). The perimeter of the rigid corner has an axis of symmetry bisecting the angle at the corner. It may be shown without difficulty that

$$a/h=\sqrt{2}-\frac{4}{\pi}, \qquad b/h=\frac{4}{\pi}-\frac{1}{\sqrt{2}}, \qquad r/h=\frac{4}{\pi}.$$

Fig. 6.



The critical width is therefore $2h(\pi+4)/\pi$ or about 4·54 times the depth of penetration. When $\theta=90°$ (flat die) calculation gave the values

$$a/h=0{\cdot}95, \qquad b/h=0{\cdot}54, \qquad r/h=5{\cdot}67.$$

The critical width is 8·6 times the width of the die. M is 3·67 die-widths below the corner. When $\theta\sim 50°$ the point H coincides with A ($\theta+\psi\sim 90°$; see fig. 2). The corresponding values of a, b and r, were calculated to be

$$a/h=0{\cdot}205, \qquad b/h=0{\cdot}51, \qquad r/h=1{\cdot}92.$$

The critical width is about 4·5 times the width of the impression. M is $2{\cdot}14h$ below the corner.

§5. Conclusions.

The minimum dimensions for the two-dimensional model are somewhat greater than those for actual indenters. For example, the minimum thickness recommended with a 136° pyramid is about twice the diagonal of the impression (Hankins and Aldous 1934). The difference is understandable since there is greater freedom of flow around a three-dimensional indentation. Moreover, the plastic zone penetrates less far, owing to the diffusion of stress concentrations through a relatively larger volume. The present calculations have, therefore, a limited practical value. The value of the investigation lies rather in its contribution to a general understanding of a hardness test, especially in showing why the hardness value is independent of specimen size, down to dimensions which are comparatively small in relation to the indentation. The investigation is also of fundamental interest in that it formulates methods for solving plane plastic problems of a kind different from any previously attempted.

Acknowledgments.

I am grateful to the Director of the British Iron and Steel Research Association for permission to publish the paper. I have received valuable help in the numerical computations from Messrs. G. Douglass and A. P. Green, of the Solid Mechanics Section of B.I.S.R.A.

References.

Field Foster, P., 1948, *Mechanical Testing of Metals and Alloys*, 4th Edition (London: Pitman & Sons), p. 161.
Hankins, G. A., and Aldous, C. W., 1934, *J. Inst. Metals*, **54**, 59.
Heyer, R. H., 1937, *Proc. Am. Soc. Test. Mat.*, **37**, 119.
Hill, R., 1947, *J. Iron & Steel Inst.*, **156**, 513; 1948, *Proc. 7th Int. Cong. App. Mech.* (London); 1949, *Quart. J. Mech. App. Maths.*, **2**, 40; 1950, *The Mathematical Theory of Plasticity*, (Oxford: Clarendon Press), Chap. IX.
Hill, R., and Lee, E. H., 1946, *Proc. 6th Int. Cong. App. Mech.* (Paris).
Hill, R., Lee, E. H., and Tupper, S. J., 1947, *Proc. Roy. Soc.* A, **188**, 273.

LXVIII. *On the Back Current in Blocking-Layer Rectifiers.*

By J. H. GISOLF,
Haarlem *.

[Received February 24, 1950.]

SUMMARY.

This paper is based on the theory of the metal semiconductor contact of Schottky and of Davidov, and is meant (*a*) to make this theory applicable in practice to all existing semiconductor types, and (*b*) to introduce quantitatively the influence of the electric field at the interface on the threshold energy. It appeared to be useful to introduce an independent variable y, proportional to the reciprocal of the concentration of the conduction electrons. A differential equation of the Abel type is derived for the electric field in the semiconductor as a function of y. Taking into account the diminution of the threshold energy by the electric field at the interface, a relation is derived between the variable y and the electric field at the interface. The differential equation is given for the different types of semiconductors. When the differential equation is solved, the current-voltage characteristic can be found by numerical integration. Some examples are given. In the case of back currents, an asymptotic solution of the differential equation exists, which leads to a parametric representation of the current-voltage characteristic in the back direction. The shape of the back current characteristic is discussed. The differences between rectifiers of the selenium type and the crystal detectors are elucidated. Some remarks on the physical background of the electrical pretreatment of selenium rectifiers are made.

§1. INTRODUCTION.

THE first theory of the blocking-layer rectifier, not based on *ad hoc* hypotheses, was given almost simultaneously by Davidov (1939) and by Schottky (1939). These theories differ in their mathematical form, but are based on the same physical conceptions (Gisolf 1942). They account satisfactorily for the current dependence of the resistance for forward and for small back currents. (The " forward " direction is the direction for which the resistance is low.) For larger back currents considerable deviations from the theory occur. The resistance first passes through a maximum and tends to vanish for large back currents. Schottky (1941) did mention the presumable cause of this deviation, namely that in consequence of the strong electric field in the semiconductor at the boundary of the metal electrode, the threshold energy for the transition of charge carriers from the metal to the semiconductor is diminished and the number of these particles in the semiconductor is increased. In the present paper a treatment of the above problem is given, which enables one to derive formulæ for the resistance of back currents with due allowance for the influence of the field on the threshold energy. The way by which this result is achieved may be outlined as follows. The concentration of the free electrons in the semiconductor is a univalent function of the distance x of the point under consideration from the surface of the electrode. Instead of x an independent variable y is introduced, which, except for a constant factor, is equal to the reciprocal of the concentration of the free electrons. From the fundamental formula of the electric field a differential equation is derived which gives the electric field strength in the semiconductor as a function of y for each value of the current density. The equation can be solved by numerical methods for the different types of semiconductors. A boundary condition is given by the field strength at a long distance from the electrode where $y=y_\infty$. Schottky's formula for the diminution of the threshold energy at the electrode as a function of the field strength together with the relation between the field strength and the variable y, leads to the value y_0 of the variable y at the electrode. The resistance and the capacity of the metal semiconductor contact are found as integrals of completely known functions of y between the limits y_0 and y_∞.

---

* Communicated by the Author.

§2. THEORETICAL PICTURE OF A SEMICONDUCTOR.

In the following pages ordinary electrons with a charge $-e$, as well as " electron holes " with a charge $+e$ will be spoken of as " electrons ". It is possible to classify the rich variety of semiconductors on account of the electron energy scheme (Gisolf 1947). The following types may be cited :—

A. The concentration of the free or conduction electrons n_c is large compared to the concentration of the electron traps N_t. The free electrons originate from impurity atoms nearly all of which are ionized by the crystal lattice. These impurities may be a stoichiometric excess of metal or metalloid atoms (for electrons and electron holes respectively), similar to those constituting the crystal lattice. The concentration n_c is then practically independent of the temperature. The conductivity slightly increases with decreasing temperature.

B. Although only a fraction of the impurity atoms is ionized by the crystal lattice, the concentration of the free electrons n_c is still large compared with the concentration of the electron traps N_t. The degree of ionization increases with increasing temperature. The concentration of the free electrons, and in consequence the conductivity, increases exponentially with increasing temperature. A definite ionization energy can be attributed to the impurity atoms. This case *B* is the only one that is treated in textbooks though it rarely occurs in practice.

C. In this case the concentration of the conduction electrons n_c is small compared to the concentration of the electron traps N_t. One can assume, without loss of general applicability, that the impurity atoms are all ionized. The trapping energies are supposed to be practically homogeneously distributed over a range of energies ΔE, which is large compared to the Boltzmann energy kT. In this case the concentration of the free electrons, and consequently the conductivity, is again an exponential function of the temperature. The pseudo "ionization energy" deduced from this exponential dependence, however, now strongly depends on the concentration of the impurity atoms N_i and in such a way that it becomes a universal linear function of the logarithm of the conductivity (the Meyer rule).

The electrons of the semiconductor, which do not belong to the atoms of the normal lattice, are either bound to impurity atoms or are "free". In any case they are supposed to have a definite energy. The energy of a free or conduction electron will be denoted by E_c and the energy of an electron bound to an impurity atom by E_i. The concentration of the conduction electrons n_c is given by the formula

$$n_c = N_c/\{1+\exp(E_c-g)/kT\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

with

$$N_c = \left(\frac{2\pi m^* kT}{h^2}\right)^{3/2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In these formulæ m^* is the "effective" electron mass. m^* may be several times the normal electron mass but the exact value cannot be fixed. T is the absolute temperature; k is Boltzmann's constant and h is Planck's constant. $N_c \gtrsim 10^{19}$ cm.$^{-3}$; g is the free energy per electron. In a stationary state the value of g is the same at every point.

The concentration n_i of the electrons bound to the impurity atoms with the energy E_i, is given by the analogous formula

$$n_i = N_i/\{1+\exp(E_i-g)/kT\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In this formula N_i is the concentration of the impurity atoms to which electrons can be bound with the energy E_i. With the abbreviation

$$y = \exp(E_c-g)/kT \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

the formulæ (1) and (3) can be written

$$n_c = N_c/(1+y) \gtrsim N_c/y, \qquad (\text{for } n_c \ll 10^{19}). \quad . \quad . \quad . \quad . \quad (5)$$

$$n_i = N_i/\{1+y\exp(-E_c+E_i)/kT\} \gtrsim N_i/y\exp(-E_c+E_i)/kT, \quad (\text{for } n_i \ll N_i) \quad . \quad . \quad . \quad (6)$$

When the contact surface of the metal and the semiconductor is flat, y only depends on the distance from the contact surface.

The space charge density ρ can readily be expressed in terms of the parameter y for each of the semiconductor types mentioned above. It is assumed that the impurity atoms in the crystal are electrically neutral so long as they have not lost electrons by thermal dissociation. For the "traps", on the other hand, it is assumed that they are electrically neutral so long as they have not captured an electron. The following formulæ for the space charge density ρ are found:

For the case A: $$\rho=e(\mathrm{N}_i-\mathrm{N}_c/y). \qquad \ldots \qquad (7)$$

For the case B: $$\rho=e\mathrm{N}_i\frac{\{y^2+y(1-\mathrm{N}_c/\mathrm{N}_i)-\mathrm{N}_c/\mathrm{N}_i\exp \Delta\mathrm{E}/k\mathrm{T}\}}{\{y^2+y(1+\exp\Delta\mathrm{E}/k\mathrm{T})+\exp\Delta\mathrm{E}/k\mathrm{T}\}}$$

$$\simeq e\frac{\{\mathrm{N}_i y^2-\mathrm{N}_c(y+\exp\Delta\mathrm{E}/k\mathrm{T})\}}{y(y+\exp\Delta\mathrm{E}/k\mathrm{T})}, \qquad \ldots \qquad (8)$$

ΔE being the dissociation energy of the impurity atoms.

For the case C: $$\rho=e\{\mathrm{N}_i-\mathrm{N}_t(1-k\mathrm{T}/\Delta\mathrm{E}\ln y)-\mathrm{N}_c/y\}, \qquad \ldots \qquad (9)$$

ΔE now being *the range* of the trapping energies of the electron traps.

The types *A*, *B* and *C* represent extreme cases. All practical cases are more or less intermediate. The bulk of the semiconductors belongs to type *C*. The discussion of this type is important for the theory of the crystal detectors. Cuprous oxide belongs most probably to type *B*. For selenium the situation is not clear. From measurements on single crystals one might conclude that selenium belongs to type *B*, with a rather small value of ΔE. For polycrystalline selenium ΔE strongly differs for variously treated samples, but it is always small. Since the value of y is large in the blocking layer, the formulæ (7) and (8) show that, when dealing with blocking layer problems, one may reckon selenium to class *A*, as has already been pointed out by Schottky (1939).

§3. DERIVATION OF THE DETECTOR EQUATION.

In the semiconductor the following two fundamental equations hold universally.

(*a*) The Poisson equation

$$d\mathrm{F}/dx=\rho(x)/\epsilon\epsilon_0. \qquad \ldots \qquad (10)$$

(*b*) The "continuity equation" for the current

$$i=n_c eu\mathrm{F}+k\mathrm{T}u\,dn_c/dx, \qquad \ldots \qquad (11)$$

where F is the electric field strength, ϵ is the dielectric constant of the semiconductor and $\epsilon_0=8{\cdot}859\,.\,10^{-14}$, A sec/V cm. is the permittivity of free space, i is the current density, e is the value of the electronic charge, u is the electron mobility in the semiconductor.

The x-axis is perpendicular to the interface of the metal and the semiconductor, the positive direction pointing to the semiconductor. By means of formula (5), n_c in equation (11) can be expressed in terms of the parameter y. A simple transformation results in

$$dy/dx=(\mathrm{F}-iy/eu\mathrm{N}_c)ey/k\mathrm{T}, \qquad \ldots \qquad (12)$$

y being a function of x only, equation (10) can be transformed in the following manner :

$$dF/dy \,.\, dy/dx = \rho(y)/\epsilon\epsilon_0. \qquad (13)$$

The function $\rho(y)$ is given by one of the formulæ (7), (8) or (9) according to the type of the semiconductor under consideration. Substituting (12) in (13) one obtains

$$dF/dy \,.\, (F - iy/euN_c)ey/kT = \rho(y)/\epsilon\epsilon_0. \qquad (14)$$

The solution of this non-linear differential equation gives F as a function of the parameter y.

Equation (14) can be shortened, by introducing an auxiliary quantity $f(y)$, defined in the following way :

$$f(y) = F(y) - iy/euN_c. \qquad (15)$$

Introducing (15) into (14), we have

$$df/dy = kT/e\epsilon\epsilon_0 \,.\, \rho(y)/yf - i/euN_c \qquad (16)$$

or

$$\boxed{df/dy = \chi(y)/f - f_c} \qquad (17)$$

where

$$\chi(y) = kT/e\epsilon\epsilon_0 \,.\, \rho(y)/y \qquad (18)$$

and

$$f_c = i/euN_c = \text{constant}. \qquad (19)$$

Equation (17) is the starting point for the crystal rectifier problem. It will be referred to as the " detector-equation ". It is a so-called Abel equation. The solutions cannot be expressed in terms of finite elementary functions. It is however very suitable for numerical integration. For each type of semiconductor the function $\chi(y)$ is completely known and can be tabulated once and for all. When the quantity $F(y)$ is found as a function of y by integration of the detector equation, it leads to the desired information concerning the electric resistance, the capacity, etc., for each value of the current density i. (See § 5.)

The quantity $f(y)$ has an obvious physical meaning. It is that part of the electric field that is needed to maintain the differences in the concentration of the free electrons from point to point.

§ 4. THE BOUNDARY CONDITIONS FOR THE SOLUTION OF THE DETECTOR EQUATION.

At a large distance from the interface, the electron distribution in the semiconductor will be the same as if no metal were present. In practice, this " large " distance means about 1 μ). According to formula (5) the concentration of the conduction electrons for $x \rightarrow \infty$ is given by

$$n_c(x \rightarrow \infty) = N_c/y_\infty, \qquad (20)$$

y_∞ is the value of y for $x = \infty$.

For $x \rightarrow \infty$ the electron distribution becomes homogeneous and in consequence $f(x \rightarrow \infty) \rightarrow 0$. Thus

$$\text{for } y \rightarrow y_\infty, \quad f(y) \rightarrow 0 \quad \text{and} \quad \chi(y) \rightarrow 0. \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Substituting (21) in the detector equation (17) one sees that for $y \rightarrow y_\infty$, df/dy becomes indefinite. We can abolish this indefiniteness by differentiation of the numerator and the denominator of the quotient χ/f. This gives for $(df/dy)_{y\infty}$ a quadratic equation with the following solutions:

$$\left(\frac{df}{dy}\right)_{y\infty} = \frac{-f_c}{2} \pm \sqrt{\frac{f^2_c}{4} + \left(\frac{d\chi}{dy}\right)_{y\infty}}, \quad . \quad . \quad . \quad . \quad . \quad (22)$$

$f(y)$ and df/dy are always negative and in (22) the sign must be chosen accordingly. The integration of the detector equation can be started at $y = y_\infty$. Because no great accuracy is wanted, the integration is readily performed.

For the case in which the current is zero, the constant f_c also becomes zero and the detector equation reduces to

$$df/dy = \chi/f \qquad (\text{for } i=0). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

This equation can be integrated immediately:

$$f^2(y) = 2 \int_{y\infty}^{y} \chi(y)\, dy \qquad \text{for } i=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

§ 5. CALCULATION OF THE CHARACTERISTIC PROPERTIES OF THE METAL-SEMICONDUCTOR CONTACT.

In the preceding section it was shown how, for a given semiconductor f can be calculated as a function of y for each value of the current density i. The next task will be to express the characteristic quantities such as the resistance, the capacity, etc., in terms of f and y.

The conductivity is given by the following formula:

$$\sigma = eun_c = eu\mathrm{N}_c/y \text{ ohm}^{-1} \text{ cm.}^{-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The barrier-layer resistance R_b is defined as the difference of the actual resistance per cm.^{2} and the resistance per cm.^{2} such as it would be if no metal were present. Thus

$$\mathrm{R}_b = \frac{1}{eu\mathrm{N}_c} \int_0^\infty (y - y_\infty)\, dx. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

From (12) it follows that

$$dx = k\mathrm{T}/e \,.\, dy/yf \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

and thus:

$$\mathrm{R}_b = \frac{k\mathrm{T}}{e^2 u\mathrm{N}_c} \int_{y_0}^{y\infty} \frac{(y - y_\infty)}{yf}\, dy \text{ ohm cm.}^2, \quad . \quad . \quad . \quad . \quad (28)$$

y_0 is the value of y at $x=0$.

Multiplying R_b by i one finds the barrier-layer tension U_b. U_b is that part of the applied voltage that appears across the barrier-layer.

$$U_b = \frac{kTi}{e^2 u N_c} \int_{y_0}^{y_\infty} \frac{(y-y_\infty)}{yf}\, dy \text{ volt.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

The electric field strength at the surface of the metal F_0 follows from (15):

$$F_0 = f(y=y_0) + iy_0/euN_c \text{ volt cm.}^{-1}. \quad . \quad . \quad . \quad . \quad . \quad (30)$$

The barrier-layer capacity C_b is given by

$$C_b = \epsilon\epsilon_0\, dF_0/dU_b \text{ farad. cm.}^{-2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

and follows therefore from (29) and (30).

§ 6. THE VALUE OF y_0 AT THE ELECTRODE.

In order to calculate the characteristic quantities of the blocking layer with the help of the formulæ of the preceding section, the value y_0 at the electrode must be known. This is synonymous with the assertion, that the concentration of the free electrons in the semiconductor at the electrode must be known.

The electrons in the semiconductor at the boundary of the metal are in thermodynamical equilibrium with the electrons inside the metal. *It is assumed that this equilibrium is not disturbed even by the largest current.* The equilibrium concentration of the free electrons at the interface is determined by the threshold energy $e\phi$ for the transition of electrons from the metal to the semiconductor. As the top of the occupied energy levels in the metal coincides with the free energy per electron g introduced in (1), this threshold energy is $kT \ln y_0$, as can be seen from (4). Thus

$$e\phi = kT \ln y_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

Now it is known that under the influence of an electric field F_0 at the boundary of the metal, the threshold energy decreases by $e\sqrt{(-eF_0/\epsilon\epsilon_0)}$. (See, for instance, *Handbuch der Physik*, xxiv., **2**, p. 437.)

This is a consequence of the " image force ". If Y_0 is the value of y at the interface in absence of any electric field, then

$$e\sqrt{(-eF_0/\epsilon\epsilon_0)} = e(\phi_0 - \phi) = kT \ln Y_0/y_0. \quad . \quad . \quad . \quad . \quad (33)$$

Eliminating F_0 from this equation with the help of formula (15) one gets

$$\boxed{f_0 + iy_0/euN_c = -k^2T^2\epsilon\epsilon_0/e^3 \,.\, \ln^2 (Y_0/y_0).} \quad . \quad . \quad . \quad (34)$$

The quantity $kT \ln Y_0$ might be called the " true " threshold energy. It is a property of the combination of the particular metal and semiconductor under consideration and must be known *a priori* in order to carry through the preceding calculations. The equation (34) describes the quantity f_0 as a function of y_0. The point of intersection of the curve representing f_0 as a function of y_0 (formula (34)) and the curve representing f as a function of y (detector equation, formula (17)) gives the values of

y_0 and f_0 which are needed to apply the formulæ (28), (29) and (30). With this result the mathematical solution of the blocking-layer problem, with due allowance for the field effect is obtained.

In the next section the method will be demonstrated with the aid of some examples. In the following section it will be shown that for large back currents an asymptotic solution of the detector equation exists which enables us to calculate the characteristic quantities of the blocking-layer without numerical integration. This asymptotic solution will then be used in the discussion of the properties of the blocking-layers at large back currents.

§ 7. SOME EXAMPLES.

In the following examples the values of the different constants, such as the dielectric constant ϵ and the electron mobility u, are not specific for a particular material, but do represent plausible values and are chosen in such a way as to make the different coefficients of the detector equation simple numbers.

For N_c we take the value 10^{-19} cm.$^{-3}$ and for u the value $u=6{\cdot}3$ cm.^{2} V^{-1} sec.$^{-1}$, which makes $euN_c=10$. For ϵ the value $\epsilon=4{\cdot}55$ is chosen and hence $kT/\epsilon\epsilon_0=10^{-8}$ V^2 cm. for $T=293°$ K.

The following examples refer to semiconductors of the types A and C. For both, the " true " threshold energy $e\phi_0$ is taken as 0·349 eV., which gives $Y_0=10^6$ at $T=293°$ K. For the semiconductor of type C the value $\varDelta E$ is arbitrarily chosen as 0·582 eV. ($=23\,kT$). The concentration of the electron donor impurity atoms and the concentration of the electron traps are chosen in such a way as to give both semiconductors at room temperature the same conductivity when no metal is present. This means that both will have the same y_∞. For y_∞ we take the value 10^4. At a large distance from the metal, that is to say for $y=y_\infty$, $\sigma=10^{-3}$ ohm cm. (see formula (25)) and the space charge denisty ρ is zero. Thus formula (7) for the semiconductor of type A gives $N_i=10^{15}$ cm.$^{-3}$ and for the semiconductor of type C formula (9) gives $N_i-0{\cdot}6N_t=10^{15}$ cm.$^{-3}$. If the arbitrary value $N_t=10^{18}$ cm.$^{-3}$ is chosen for N_t, it follows that $N_i=6{\cdot}01\;\;10^{17}$ cm.$^{-3}$. The numerical integration of the equation (17) has been performed by means of a very straightforward method, due to Horst van Sanden (1946). The integrations are carried through for a number of positive and negative values of i.

In the Tables I. and II. some of the results are summarized with the purpose of demonstrating that two semiconductors, which are alike from the electrical point of view, may behave quite differently as rectifiers in contact with the same metal even when the threshold energy is also the same. For the type C semiconductor the blocking-layer resistance is much less than for the type A semiconductor and for small current densities the direction of the rectification is opposite in the two examples, as can be seen from the current-voltage characteristics in fig. 1. These differences are brought about by the much larger space charge density in the case of the type C semiconductor and its strong current dependence.

TABLE I.

Type A : $Y_0=10^6$; $N_i=10^{15}$ cm.$^{-3}$.

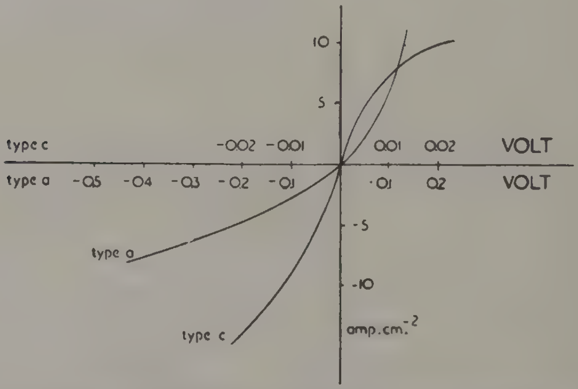
i A cm.$^{-2}$	R Ohm cm.^{2}	y_0	F_0 V. cm.$^{-1}$	ρ_0 el. charges	$\phi_0-\phi$ volt
+10	$1{\cdot}08\times10^{-2}$	$1{\cdot}00\times10^6$	$+9{\cdot}02\times10^3$	$9{\cdot}9\times10^{14}$	0
+1	$4{\cdot}36\times10^{-2}$	$2{\cdot}45\times10^5$	$-3{\cdot}20\times10^3$	$9{\cdot}9\times10^{14}$	0·036
0	$6{\cdot}40\times10^{-2}$	$1{\cdot}48\times10^5$	$-5{\cdot}93\times10^3$	$9{\cdot}3\times10^{14}$	0·048
−1	$8{\cdot}10\times10^{-2}$	$8{\cdot}06\times10^4$	$-9{\cdot}80\times10^3$	$8{\cdot}8\times10^{14}$	0·064
−10	$5{\cdot}40\times10^{-2}$	$2{\cdot}28\times10^4$	$-2{\cdot}31\times10^4$	$5{\cdot}6\times10^{14}$	0·095

TABLE II.

Type C : $Y_0=10^6$; $N_i=6{\cdot}10^{17}$ cm.$^{-3}$; $N_t=10^{18}$ cm.$^{-3}$; $\Delta E=0{\cdot}582$ eV.

i A cm.$^{-2}$	R Ohm cm.^{2}	y_0	F_0 V. cm.$^{-1}$	ρ_0 el. charges	$\phi_0-\phi$ volt
+10	$2{\cdot}38\times10^{-3}$	$5{\cdot}70\times10^4$	$-1{\cdot}34\times10^4$	$7{\cdot}54\times10^{16}$	0·073
+1	$1{\cdot}65\times10^{-3}$	$2{\cdot}93\times10^4$	$-2{\cdot}02\times10^4$	$4{\cdot}64\times10^{16}$	0·089
0	$1{\cdot}60\times10^{-3}$	$2{\cdot}76\times10^4$	$-2{\cdot}08\times10^4$	$4{\cdot}36\times10^{16}$	0·091
−1	$1{\cdot}53\times10^{-3}$	$2{\cdot}60\times10^4$	$-2{\cdot}15\times10^4$	$4{\cdot}11\times10^{16}$	0·094
−10	$1{\cdot}16\times10^{-3}$	$1{\cdot}80\times10^4$	$-2{\cdot}62\times10^4$	$2{\cdot}49\times10^{16}$	0·101

Fig. 1.



The barrier-layer tension as a function of the current density of two crystal-rectifiers of the A-type and the C-type respectively.
For the A-type : $Y_0=10^6$; $N_i=10^{15}$ cm.$^{-3}$.
For the C-type : $Y_0=10^6$; $N_i=6{\cdot}10^{17}$ cm.$^{-3}$; $N_t=10^{18}$ cm^{-3} ; $\Delta E=0{\cdot}582$ eV.

§ 8. AN ASYMPTOTICAL SOLUTION OF THE ABEL EQUATION.

Consider the case in which the constant f_c in the differential equation (17) is negative. Physically this means that a back-current flows. When y is increased to infinity, the function $\chi=\rho(y)/y$, tends to zero as can be seen from the formulæ (7), (8) and (9). f being always negative, df/dy cannot tend to a finite positive value and therefore χ/f cannot become zero for large values of y. But then f and df/dy must go to zero with increasing y and thus the differential equation reduces to

$$f(y)=\chi(y)/f_c. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Substituting (35) into (28) and considering that $y \gg y_\infty$, one finds for the back-resistance the approximation

$$R_b = \frac{kT}{e} \frac{i}{(euN_c)^2} \int_{y_0}^{y_\infty} \frac{dy}{\chi(y)} \qquad (36)$$

The above approximation holds as well for large values of y, as for large negative values of i. In this approximation the equation (34) reduces to the following transcendental equation for y_0:

$$iy_0/euN_c = -k^2T^2\epsilon\epsilon_0/e^3 \,.\, \ln^2 (Y_0/y_0). \qquad (37)$$

In the next section the above approximation will be used to calculate the back-current characteristic of the contact of a metal with a semiconductor of the A-type.

§ 9. THE BACK-CURRENT CHARACTERISTIC FOR A SEMICONDUCTOR OF THE A-TYPE.

For a semiconductor of the A-type, equation (35) reduces to

$$f = \frac{kTeuN_cN_i}{\epsilon\epsilon_0 i} \cdot \frac{(y-y_\infty)}{y^2} \qquad (38)$$

and (29) reduces to

$$U_b = -\frac{\epsilon\epsilon_0}{2eN_i(euN_c)^2} \cdot i^2(y_0^2 - y_\infty^2). \qquad (39)$$

From equations (37) and (39) the equation for the back-current characteristic can be derived, computing y_0 from (39) and substituting in (37). For practical calculations, however, it is more convenient to consider equations (37) and (39) as a parameter representation of the back-current characteristic. For each chosen value of y_0 these equations give the corresponding values of i and U_b.

When the back-current increases, the value of y_0 decreases, as can be seen when computing di/dy from (37). Thus in (39) the factor $(y_0^2 - y_\infty^2)$ will decrease when i^2 increases and the possibility of a maximum in the border-layer tension exists. With the aid of (37), i can be eliminated from (39). For the occurrence of a maximum for U_b dU_b/dy_0 must be zero. This leads to the following transcendental equation for y_0^*, the value of y_0 for which the barrier-layer back tension attains its maximum:

$$\ln Y_0/y_0^* = 2(y_0^* - y_\infty^2)/y_\infty^2. \qquad (40)$$

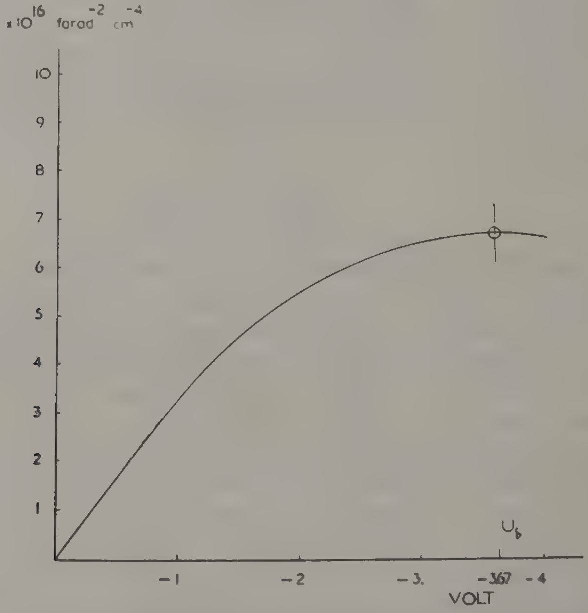
Substituting the value of y_0^* found from this equation into (37) one finds the corresponding value of the back-current i^*, and substitution of y_0^* and i^* into (39) gives the maximum U_b^* of the barrier-layer tension. Equation (40) can be solved graphically in a simple way.

Some examples:

	N_i	Y_0	y_0^*	U_b^*	i^*
(*a*)	10^{15} cm.$^{-3}$	10^7	$2{\cdot}03 \times 10^4$	$-3{\cdot}71$ volt	-30 A/cm.^{2}
(*b*)	10^{15} cm.$^{-3}$	$1{\cdot}4 \times 10^{10}$	$2{\cdot}62 \times 10^4$	-100 volt	-107 A/cm.^{2}
(*c*)	10^{15} cm.$^{-3}$	10^{12}	$3{\cdot}07 \times 10^4$	-279 volt	-157 A/cm.^{2}

The voltage maximum in the back direction cannot be observed on selenium or cuprox rectifiers because at current densities much smaller than those at which the maximum would occur, the rectifier would have been disturbed by heat. The maximum is observable on crystal rectifiers when high-frequency voltage pulses are applied (Torrey and Whitmer 1948). The maximum is found at a voltage of some 100 or 200 volts. The value of the current density in crystal rectifiers of the whisker-contact type is difficult to estimate, because the contact area between the whisker and the crystal is not exactly known. At any rate, the back-current densities at high back voltages can be estimated at magnitudes of a hundred A/cm.^{2} or so.

Fig. 2.



$1/C_b^2$ of a crystal rectifier of the A-type as a function of the back-voltage U_b. $Y_0=10^7$; $N_t=10^{15}$ cm.$^{-3}$.

The example (*a*) corresponds to a rather bad rectifier ; the example (*b*) corresponds to a modern high-inverse-voltage crystal detector and the example (*c*) corresponds to a rectifier of the selenium type.

The *barrier-layer capacity* C_b follows from (31) :

$$dF_0/dU_b=(dF_0/dy_0)/(dU_b/dy_0), \quad \ldots \ldots \quad (41)$$

dF_0/dy_0 follows immediately from (33) and dU_b/dy_0 follows from (39), when i is eliminated with the aid of (37).

As an example $1/C_b^2$ is calculated as a function of U_b for a semiconductor of the A-type with $N_i=10^{15}$ cm.$^{-3}$ and $Y_0=10^7$, (see fig. 2).

For small back-voltages $1/C_b^2$ is the same linear function of the back-voltage as has been found by Schottky (1941–1942).

For large back-voltages deviations from linearity are observed which agree qualitatively with the calculated curve of fig. 2. The curve has a maximum analogous to the maximum in the current-voltage characteristic in the back direction.

The maximum occurs at a back-voltage U_b^{**} corresponding to a value y_0^{**} in the parameter representation of (37) and (39), *i. e.* to the value which is a solution of the transcendental equation

$$\ln Y_0/y_0^{**}=\tfrac{4}{3}(y_0^{2**}-y_\infty^2)/y_\infty^2 . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (42)$$

In the example of fig. 2 one finds $y_0^{**}=2\cdot35\times10^4$, and substituting this value into (37) one finds $i^{**}=-25\cdot3$ A/cm.$^{-2}$. Substituting y_0^{**} and i^{**} into (39) one finds $U_b^{**}=-3\cdot67$ V. Just as with the maximum in the current-voltage characteristic it is to be expected that the maximum in the $(1/C_b^2, U_b)$ curve will only be observable with short voltage-pulses. As far as the author is aware, such a maximum has not yet been observed †.

§ 10. THE RECTIFIER OF THE SELENIUM-TYPE.

Among the crystal rectifiers used in practical applications two quite different types can be discerned: the selenium and the cuprous oxide rectifiers on the one side and crystal detectors on the other. The principal difference lies in the fact that the crystal detector operates at 10^4 or 10^5 times as large a current density as the selenium or the cuprous oxide rectifier. That the applied current density is of preponderant influence on the unipolar conductivity of a metal semiconductor contact, will be clear when comparing, for example, fig. 1 and fig. 3. Most semiconductors in contact with a fine metal point, have an asymmetrical current-voltage characteristic, when the current density is made sufficiently high. This is effected by touching the crystal very loosely, so that the contact area is very small. Extremely few semiconductors have good rectifying properties at as low a current density as about a milliampere per cm.^{2}. Only two are used in practice, viz. selenium and cuprous oxide.

For a good rectifier of the selenium or cuprous oxide type, it is claimed that at a back-voltage of about 10 volt, the back-current will not exceed one milliampere per cm.^{2}. The contact being characterized by the two parameters Y_0 and N_i, one can ask for which pairs of values of these parameters a metal semiconductor contact will satisfy this requirement. In the formulæ (37) and (39) we substitute for i the value -10^{-3} amp cm.$^{-2}$ and for U_b the value -10 volt and thereupon these equations are solved for Y_0 and N_i. This is performed by putting N_i arbitrarily equal to integral powers of ten and computing the corresponding values of Y_0. The pairs of values thus obtained, together with some other calculated features of the contact, are tabulated in Table III. Columns 1 and 2 give N_i and Y_0. Column 3 gives the value of ϕ_0 corresponding to Y_0. Column 4 gives the value of y_0. Column 5 gives the electric field at the electrode. In column 6 V_d the so-called "diffusion-potential" is tabulated, as derived from

---

† *Note added in proof.*—R. N. Smith (1946, *Phys. Rev.*, **69**, 683), reports that the contact capacity of Si-detectors, plotted as a function of the voltage, shows a minimum at a backvoltage of 0·5 volt.

$$V_d = kT \ln Y_0/y_\infty = 0{\cdot}058 \log Y_0/y_\infty \text{ volt.} \quad . \quad . \quad . \quad . \quad (43)$$

Column 7 gives the conductivity σ of the semiconductor.

It must be kept in mind, that the figures in this table are not calculated with values of ϵ, u and m^* which hold exactly for selenium; these figures are therefore only meant as an illustration. The quantities that are open to experimental check are: N_i (from the voltage dependence of the capacity); V_d (from the current-voltage characteristic in the forward-direction); and σ (also from the current-voltage characteristic in the forward-direction). The figures inside the framed part of Table III. agree with experimental observations on actual rectifiers.

TABLE III.

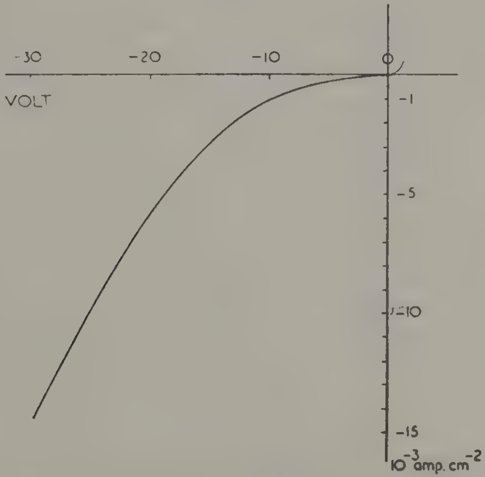
N_i in cm.$^{-3}$	Y_0	ϕ_0 in volt	y_0	F_0 in volt cm.$^{-1}$	V_d in volt	σ in Ohm^{-1} cm.$^{-1}$
10^{18}	$4{\cdot}26\times10^{28}$	1·66	$2{\cdot}81\times10^{10}$	$2{\cdot}81\times10^{6}$	1·61	1
10^{17}	$1{\cdot}39\times10^{20}$	1·17	$8{\cdot}88\times10^{9}$	$8{\cdot}88\times10^{5}$	1·05	10
10^{16}	$1{\cdot}53\times10^{14}$	0·82	$2{\cdot}81\times10^{9}$	$2{\cdot}81\times10^{5}$	0·65	10^2
10^{15}	$1{\cdot}47\times10^{12}$	0·71	$8{\cdot}88\times10^{8}$	$8{\cdot}88\times10^{4}$	0·48	10^3
10^{14}	$1{\cdot}84\times10^{10}$	0·60	$2{\cdot}81\times10^{8}$	$2{\cdot}81\times10^{4}$	0·31	10^4
10^{13}	$9{\cdot}24\times10^{8}$	0·52	$8{\cdot}88\times10^{7}$	$8{\cdot}88\times10^{3}$	0·17	10^5
10^{12}	$1{\cdot}16\times10^{8}$	0·47	$2{\cdot}98\times10^{7}$	$2{\cdot}98\times10^{3}$	0·062	10^6
10^{11}	$1{\cdot}46\times10^{9}$	0·53	$1{\cdot}09\times10^{8}$	$1{\cdot}09\times10^{4}$	0·068	10^7

In fig. 3 the back-current characteristic is drawn for a rather arbitrary example, as calculated according to § 9. The decreasing back-resistance with increasing back-voltage is a common experience on selenium rectifiers.

Surveying Table III. the most striking feature is the small range for the threshold potential ϕ_0 in which rectifiers are actually realized. This is the more astonishing, as good selenium rectifiers can be made with a rather large variety of electrode metals. Now, it is known that every selenium rectifier gets its rectifying properties not until it has been subjected to an electrical treatment. Usually a constant back-current is forced through the rectifier, the back-voltage increasing in the course of the treatment until the desired back-voltage is reached. The question arises as to which of the two characteristic parameters Y_0 and N_i is affected by this treatment. It has been suggested (F.I.A.T. Report), that during the electrical treatment under the influence of the electric field the impurity atoms, which are assumed to be electrically negative, are driven from the semiconductor to the metal. There they should either be stored in a very thin layer at the interface, or be taken up by the metal, but in all events to be put out of action. A reduction of the impurity atoms will increase the back resistance.

The curve of fig. 4 represents the back-voltage for a back-current of 10^{-3} A/cm.$^{-2}$, as a function of the concentration of the impurity atoms at an arbitrarily chosen value of Y_0. As can be seen from this curve, which would describe the falling off of the concentration N_i during the electrical treatment, the concentration of the impurity atoms must be diminished by a factor $\sim 10^2$, to compensate a " mismatch " in the threshold potential ϕ_0 of $\sim 10^{-1}$ volt. Curves as of fig. 4 have been calculated for very different values of Y_0, leading to practically the same result. This result makes it rather improbable that the effect of the electrical pretreatment is due to a diminution of the impurity atoms. The scanty experimental data concerning the change of N_i during the electrical treatment show a decrease of N_i during the last stage of the electrical treatment of a factor 5 at most.

Fig. 3.

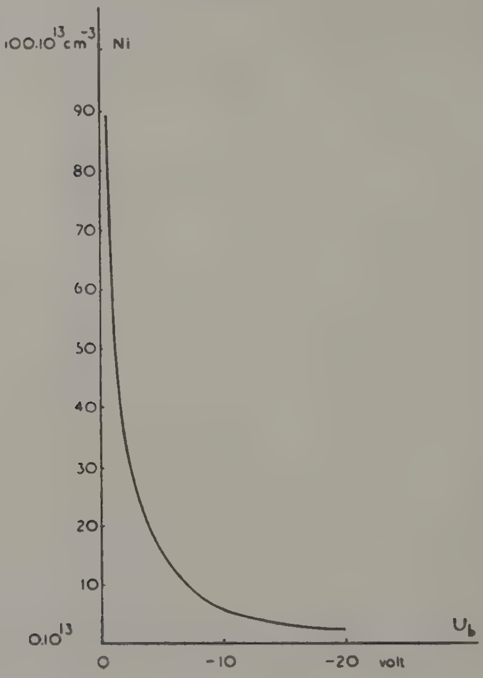


The barrier-layer tension U_b as a function of the back-current of a crystal rectifier of the A-type. $N_i = 10^{15}$ cm.$^{-3}$; $Y_0 = 1{\cdot}47 \times 10^{12}$. This example is meant to represent the behaviour of a normal Selenium rectifier.

It is conceivable that the observed influence of the electrical pretreatment on the concentration of the impurity atoms is only a secondary effect and that the primary effect is an increase of the threshold potential ϕ_0. One might for instance suppose that during the treatment under the influence of the electric field at the boundary of the metal, metal ions enter the selenium, building up a double layer at the interface, by which the threshold potential ϕ_0 for electron holes is increased. In this assumption the curve of fig. 5 is calculated for $N_i = 10^{15}$ cm.$^{-3}$ and $i = 10^{-3}$ A/cm.$^{-2}$. It appears that an increase of ϕ_0 of 0·1 volt only, can raise the back-voltage to an enormous extent. It is from present day experimental knowledge not possible to decide between the two suggestions. In virtue of the foregoing considerations the author attaches the greatest probability to the last.

From Table III. it is evident that the chance that an arbitrary combination of a metal and a semiconductor will be a good rectifier, is extremely small, and so is the number of the actual technically useful combinations. The copper-cuprous-oxide rectifier is a case of fortunate coincidence, very sensitive to traces of admixtures and the thermal pretreatment. The great technical success of the selenium rectifier is due to the possibility of adapting it to its purpose by the simple electrical pretreatment, and this possibility is linked up with the extraordinary mobility of foreign ions in hexagonal selenium (Torrey and Whitmer 1948). But this merit is also its weakness, for the electric fields during the normal operation of the rectifier give rise to the same effect as those during the electrical pretreatment and this results in a sometimes very appreciably increasing resistance both in the forward and in the back direction.

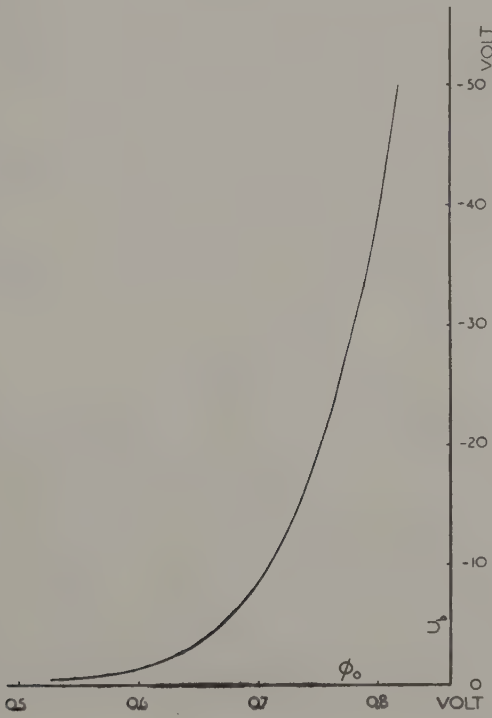
Fig. 4.



The back-voltage at a current density of 10^{-3} A × cm.$^{-2}$ as a function of the concentration of the impurity atoms N_i for a rectifier of the A-type, with $Y_0=10^{10}$.

In connection with the above considerations the effect of artificial barrier-layers, which are sometimes applied, for instance in the form of a thin film of lacquer between the selenium and the metal, can easily be explained. When the thin lacquer film is pierced at several spots, the local current density in the different parallel current paths is much higher than it would have been, when the same current was homogeneously distributed over the whole cross-section of the rectifier. It follows from the shape of the current-voltage characteristic that the rectifier-effect is strongly favoured by an increase of the current density. The effect of the pierced insulating film is thus essentially the same as that of a loosely applied whisker contact.

Fig. 5.



The back-voltage at a current density of 10^{-3} A × cm.$^{-2}$ as a function of the threshold potential ϕ_0 for a rectifier of the A-type with a concentration of the impurity atoms N_i of 10^{15} cm.$^{-3}$.

References.

Davidov, B., 1939, *J. Phys., U.S.S.R.*, **1,** 167.
F.I.A.T. Report on Selenium Dry Rectifier Development. Final Report No 706.
Gisolf, J. H., 1942, *Ned. T. Natuurk.*, **9,** 49 ; 1947, *Ann. Physik* 6 *Folge*, **1,** 3.
Sanden, Horst van, 1946, *Praxis der Differentialgleichungen.*
Schottky, W., and Spenke, E., 1939, *Wiss. Veröff Siemens-Werke*, **18,** 225.
Schottky, W., 1941–1942, *Z. Phys.*, **118,** 539.
Torrey, H., and Whitmer, Ch., 1948, *Crystal Rectifiers*, Chapter 12, New York.

LXIX. *The Electrical Conductivity of Bismuth Fibres.—* II. *Anomalies in the Magneto-Resistance.*

By B. Donovan and G. K. T. Conn,
Department of Physics, The University of Sheffield*.

[Received May 19, 1950.]

Summary.

In examining the magneto-resistance of thin fibres of bismuth, which are known from previous work to approximate to single crystals, it was found that a considerable proportion of these, about 10 per cent, showed a decrease in the resistance on application of a magnetic field. The effect is only present in fields of the order of 1000 oersteds or less and is of the order of $10^{-4}\rho_0$. The influence of direction of current, direction of field and orientation of the crystal axes is examined. It is considered that these observations provide the first unequivocal evidence of the existence of a longitudinal Hall coefficient. Used in conjunction with the theory of conduction in anisotropic solids, it is possible to indicate the appropriate crystal orientation. It has not, as yet, been possible to designate this with accuracy.

§ 1. Introduction.

In the first paper of this series (Donovan and Conn 1949), the initial results of an investigation into the magneto-resistive behaviour of bismuth fibres were discussed, and particular attention was paid to the elucidation of the structure of the specimens. Comparison of the results with those obtained for single crystals (Stierstadt 1933) led to the conclusion that the fibres were, with certain exceptions, single crystals with the main cleavage plane parallel to the fibre axis. The exceptions were provided by the low-resistance (*i.e.* the thicker) fibres and it seemed feasible that these should possess a less highly developed crystal structure; their magneto-resistive behaviour was not radically different from that of the high-resistance fibres, but was sufficiently distinctive to enable positive identification to be made.

In the course of this investigation certain anomalous results were obtained and it is the purpose of the present paper to discuss these. The essential feature is that, under certain circumstances, the change in resistance resulting from the application of the magnetic field may be negative, or, in other words, the change in resistance may reverse in sign when the direction of the magnetic field is reversed. The possibility of such an "Umkehreffekt" has been discussed in a number of theoretical papers, but no conclusive experimental evidence has hitherto been reported. The work of Kapitza

---

* Communicated by Professor W. Sucksmith, F.R.S.

(1928) and others has lead to the conclusion that the change in resistance is proportional to the square of the field strength in weak fields and tends gradually to a linear relation at very high field strengths. The results discussed below indicate that the resistance cannot be represented simply as a quadratic function of the field strength and suggest that a linear term is also present which outweighs the quadratic term below a certain critical field strength. A preliminary notification of the present work has already appeared (Conn and Donovan 1948).

This paper does not contain a detailed treatment of the "Umkehreffekt" as such, but presents experimental results which may be regarded as evidence of its existence in bismuth; in the absence of sufficiently extensive experimental data, a complete interpretation cannot be attempted, but it is felt that a qualitative explanation can with advantage be given.

§ 2. EXPERIMENTAL RESULTS.

The preparation and mounting of the fibres, and the experimental technique employed, have been described in the previous paper (Donovan and Conn, *loc. cit.*). The same potentiometer arrangement was used in the present series of measurements and it was again found convenient to use the galvanometer as a continuous recorder of potential difference. In the earlier experiments, the change in resistance of the specimen was determined as a function of its orientation with respect to the magnetic field, which remained constant for a given " cycle." In the experiments to be described the fibre was maintained in a fixed position with respect to the field and its resistance was obtained as a function of the field strength

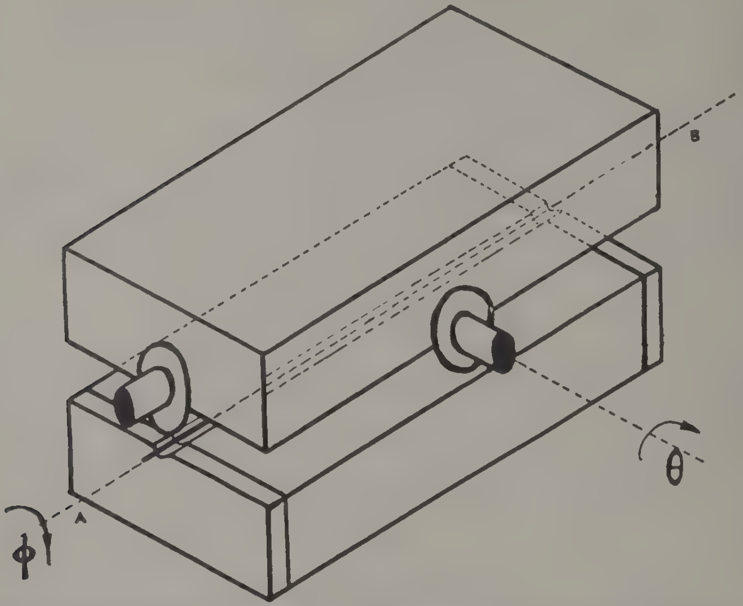
At low field strengths the residual magnetism of the pole-pieces presents difficulties and it was deemed advisable to employ a source of magnetic field which did not involve the use of iron cores. Two large coils were available, each wound with 1790 turns, the outer diameter being 36 cm. and the inner diameter 12 cm. These were supported coaxially, as in an Helmholtz arrangement. With a separation of 3·8 cm. a field of 120 oersteds per ampere was obtained at the centre of the gap. Owing to the small dimensions of the specimens, no difficulties arising from inhomogeneities of the field were encountered.

With this arrangement the procedure of varying the field was considerably simplified.

It will be convenient to refer to the ϕ and θ orientations already introduced (Donovan and Conn, *loc. cit.*) and fig. 1 has been reproduced to make the notation clear. In both cases the axis of rotation is perpendicular to the magnetic field; the axis of the fibre is parallel to, and perpendicular to the axis of rotation in the ϕ and θ cases respectively. In the latter case, therefore, the fibre rotates in a plane parallel to the magnetic field.

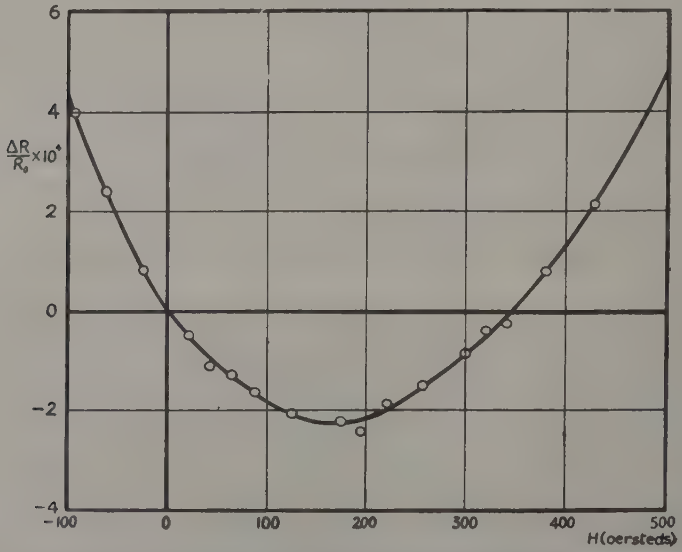
In the ϕ case the anomalous behaviour has appeared in certain fibres and is illustrated in fig. 2 which shows the change of resistance plotted against field strength for a fibre whose orientation corresponds to the minimum of the ϕ curve. The fibre had a zero-field resistance of 170 ohms. It will be seen that for field strengths less than approximately 350 oersteds the change in resistance is negative and the minimum value occurs in the neighbourhood of 180 oersteds. The decrease in resistance is extremely small but,

Fig. 1.



Isometric drawing of the holder. AB is the axis of the fibre and the ϕ-axis.

Fig. 2.



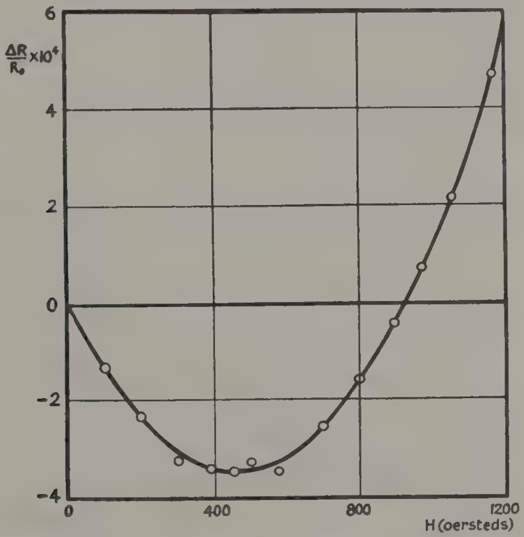
Variation of resistance with field-strength for 170 ohm fibre.

by virtue of the comparatively high resistance of the specimen and the sensitivity of the arrangement, it is easily detected and is readily reproducible.

It should be emphasized that this phenomenon is not characteristic of bismuth fibres *per seq*, and in fact only a fraction of the order of 10 per cent of the fibres exhibited this effect. For this minority the effect was quickly and convincingly demonstrated. Having balanced the potentiometer in the absence of a magnetic field, a suitable permanent magnet is brought up to the specimen and the galvanometer observed.

All the fibres which showed the negative change in resistance did so when the orientation corresponded to the minimum of the ϕ curve. With one exception, which will be referred to later, no such effect was observed when the orientation corresponded to the maximum of the ϕ curve. The significance of these two extreme positions lies in the fact that they represent simple orientations of the main cleavage plane with respect to the field (Stierstadt 1933, Donovan and Conn *loc. cit.*). No negative change in resistance was observed in any fibre when its axis was parallel to the field (θ-orientation).

Fig. 3.



Variation of resistance with field-strength for 6 ohm fibre.

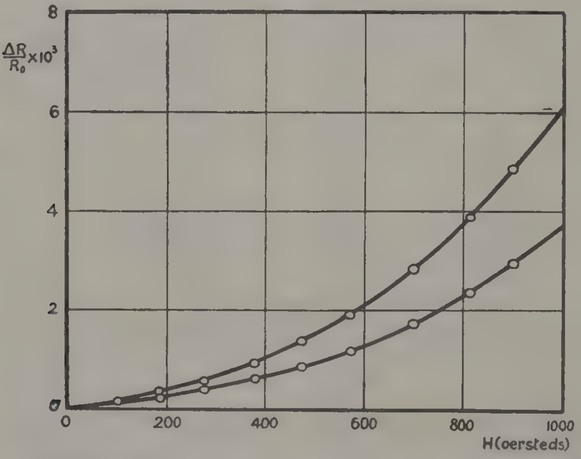
Considerable variation was apparent in the order of magnitude associated with the anomaly. In particular, the range of field strength over which the resistance change was negative varied widely from one specimen to another; the limits of the critical or "change-over" field strength were roughly 300 to 1000 oersteds. No useful purpose will be served by including a large number of experimental curves, but a further example may be given (fig. 3) illustrating a wider range of negative resistance change.

The critical field strength in this case is approximately 930 oersteds. The fibre had a zero-field resistance of 6 ohms and again was set at the minimum of the ϕ curve.

The magneto-resistive behaviour of the fibres which did not exhibit the effect in question was quite consistent, and is illustrated in fig. 4, which was obtained with a representative "orthodox" specimen. The change of resistance for weak fields is shown for orientations corresponding to the maximum and minimum of the ϕ curve

Finally, we give curves (fig 5) obtained with a fibre which was unique in its behaviour and which provided the solitary exception previously mentioned. This fibre had a zero-field resistance of 46 ohms and was remarkable in two respects: (*a*) the decrease in resistance was about fifty times larger than that obtained in other cases and (*b*) the decrease in resistance was obtained with orientations corresponding to the maximum as well as the minimum of the ϕ curve. The critical field strengths in these two cases are approximately 1300 and 1700 oersteds respectively, and these are considerably greater than any other observed values.

Fig. 4.



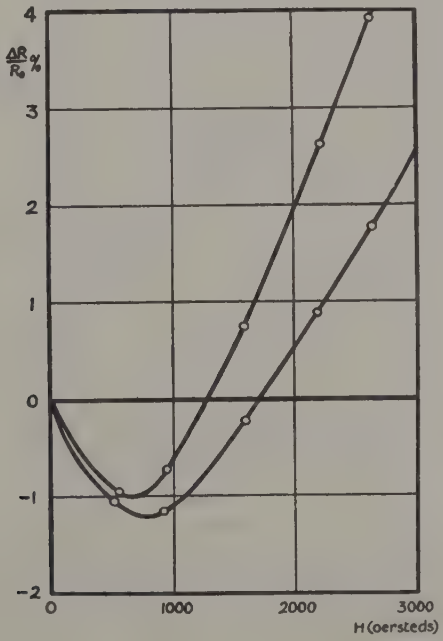
"Orthodox" magneto-resistance effect in weak fields. The upper and lower curves represent the maximum and minimum respectively of the ϕ cycle.

§ 3. POSSIBLE ERRORS.

In view of the small magnitude of the decrease in resistance, and the nature of the phenomenon, suspicion is at first to be directed towards possible sources of experimental error. The various effects, however, which might conceivably give rise to spurious results can be discussed, if not with, finality, at least with sufficient cogency to enable the "Umkehreffekt" to be regarded as genuine.

The possible sources of error which are relevant in magneto-resistance measurements have been enumerated by Thompson (1938). Of these, the most serious is undoubtedly due to the presence of the transverse Hall effect. If the electrodes are attached to the specimen in such a way that they do not lie on a line of current flow, then, when the magnetic field is applied, a component of the Hall e.m.f. will appear across the terminals. Such a misalignment is illustrated, in exaggerated form, in fig. 6 (*a*) which represents a section through a specimen perpendicular to the magnetic field. The Hall e.m.f. reverses, of course, when the field is reversed and the majority of previous workers have used mean values obtained from the two directions of the field. Under these circumstances the resistance change must necessarily be a symmetrical function of the field and no " Umkehreffekt " could have been detected. Moreover, it should be pointed out that the decrease in resistance is associated essentially with weak fields and that, in the past, more attention has been paid to investigations in strong fields, or at low temperatures, where the resistance changes are very large.

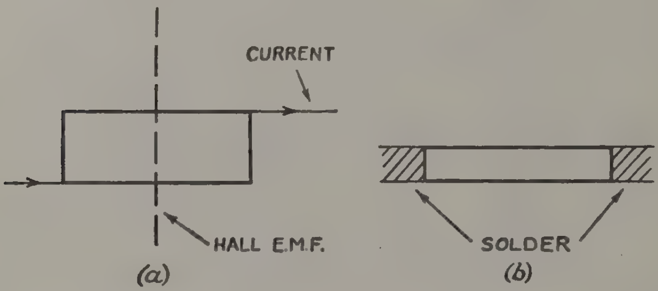
Fig. 5.



Variation of resistance with field-strength for 46 ohm fibre.

As mentioned in the previous paper already referred to, specimens in the form of fibres possess the distinctive advantage of combining the structure of a single crystal with a resistance value suitably high for accurate measurements. In addition, the shape of the specimens is such as to minimize the influence of transverse effects, and in this respect fibres are superior to the previous types of single crystal, such as triangular prisms (Stierstadt 1933) and plates (Thompson 1938). Two other reasons may be adduced for believing that the Hall effect referred to above is not responsible for the observed negative resistance change. Firstly, the electrodes are not " localized " at points on the specimen, as would be the case for instance with fine wires; the method of construction ensures that contact is made over the whole cross-section, as shown in fig. 6 (*b*). Thus any effects due to the type of misalignment depicted in fig. 6 (*a*) are eliminated. Secondly, the effect is not universal and it is difficult to see how such a general explanation as the Hall effect can be valid when about 90 per cent of the fibres show no indication of the phenomenon.

Fig. 6.



(*a*) Possible misalignment of specimen. Magnetic field perpendicular to plane of paper.
(*b*) Type of mounting used for fibres.

The evidence indicates an explanation in terms of the crystal structure of the specimen, and suggests that the " anomalous " fibres are characterized by departure from the general structure discussed in the previous paper. Whereas in the majority of fibres the main cleavage plane is parallel to the fibre axis, it seems reasonable to suppose that in the " anomalous " cases this no longer applies, and the main cleavage plane is inclined at a small angle to the axis. Small variations in this angle might well account for the different values of the " change-over " field strength observed.

It is natural to seek confirmation of such conclusions by examining the X-ray diffraction of these fibres. A preliminary investigation undertaken through the kind co-operation of Mr. R. B. Shepherd has served to clarify the problem but it cannot be said as yet to have yielded clear evidence. The inclination of the cleavage plane to the axis, upon which our interpretation is based, is sufficiently small to make discrimination difficult by X-ray methods and, bearing in mind that the fibres are of the order of 10^{-3} cm. in diameter, it would appear that this corroboration, which is clearly desirable, is likely to require a separate investigation in itself.

§ 4. OUTLINE OF PREVIOUS WORK.

Before any discussion of the foregoing experimental results is attempted, it is necessary to give a brief survey of the existing literature. The theoretical approach may be directed along one of two lines: the starting point may be either the classical equations of conduction in an anisotropic medium, or the quantum mechanical formulation of transport phenomena. The latter method of treatment has not, hitherto, proved particularly satisfactory in the case of bismuth, as the effect of the crystal symmetry has not been rigorously investigated.

The negative change in resistance would be readily explicable if the Hall effect contained a longitudinal as well as a transverse component, *i. e.* if an e.m.f. were produced *parallel* to the current flow. This must not be confused with the spurious Hall effect due to a bad experimental arrangement considered above. Such a longitudinal Hall effect would give rise to a resistance change whose sign depends on the direction of the magnetic field. The transverse character of the Hall effect was postulated by Voigt (1928) but the consequences of removing the restriction were examined by Kohler (1934) following some experimental results on the Hall effect (Verleger 1932) which could not be reconciled with Voigt's treatment. In view of the considerable experimental evidence supporting the Voigt theory, Shoenberg (1935) was led to suggest that Verleger had incorrectly described his experiments and that in fact the discrepancy was not genuine. Shoenberg dealt with some of the implications of Kohler's work and endeavoured to obtain experimental evidence by means of a magneto-resistance experiment with a bismuth crystal. No indication, however, of a longitudinal Hall effect was forthcoming.

A more detailed theoretical investigation of the magneto-resistance effect, based on Kohler's work, was carried out by Meixner (1939). Starting with the general expression for the resistance, in terms of odd as well as even powers of the field strength, Meixner showed that an "Umkehreffekt" could occur in anisotropic substances, for certain orientations of the current, field and crystal axes. In bismuth, when the conditions for maximum "Umkehreffekt" are fulfilled, the linear term should out-weigh the quadratic term for field strengths less than about 6700 oersteds and the change of resistance should be negative.

In a later paper Meixner (1941) made a comprehensive examination of electrical transport phenomena, from the quantum mechanical standpoint. This treatment, which was quite general, led to the result that neither the electrical nor the thermal conductivity was dependent upon the direction of the magnetic field. Any "Umkehreffekt" was thus inexplicable on this basis, but, in surveying the experimental results on the Hall effect in bismuth, Meixner admitted that the question must be regarded as still open. Some significance is to be attached to the fact that an "Umkehreffekt" of thermal conductivity, also excluded by Meixner's theory, has been observed in bismuth by Grüneisen and Gielessen (1936). It would appear that the quantum mechanical formulation necessarily involves assumptions that render it invalid in the case of bismuth. In a further application of the electron theory of metals, Kohler (1941) suggested that a possible explanation of these discrepancies might lie in the neglect of the quantization of the electron paths in the magnetic field.

The only previous experimental evidence of this effect for bismuth has been reported by Casimir and Gerritsen (1941). Their measurement showed that the resistance was actually a function of the direction of the field, for a single crystal at 77·3° K. Casimir and Gerritsen concluded, however, that the cause was spurious, since a similar "Umkehreffekt" was found with a crystal orientation which was precluded on theoretical grounds. No reference was made to Shoenberg's paper and, as no details of the experimental arrangement were given, the interpretation of the results cannot profitably be discussed.

§ 5. DISCUSSION OF RESULTS.

We examine now the implications of the relevant theoretical results (Shoenberg 1935) for various special cases and conclude by advancing a tentative explanation to account for the results in the second section of this paper.

The most general expression for the Hall effect is

$$E_i = R_{ikl} J_k H_l, \qquad \ldots\ldots\ldots \qquad \text{(i)}$$

where **E**, **J**, **H** are vectors denoting the electric field responsible for the Hall e.m.f., the current density and the magnetic field respectively. The R_{ikl} are governed by the symmetry of the crystal and in the case of bismuth it has been shown by Kohler (1934) that there are only four independent constants:

$$\left.\begin{aligned} &R_{231} = -R_{132} = R_1, \\ &R_{312} = -R_{321} = R_2, \\ &R_{123} = -R_{213} = R_3, \\ &R_{111} = -R_{122} = -R_{212} = -R_{221} = R_4. \end{aligned}\right\} \qquad \ldots\ldots \qquad \text{(ii)}$$

If there is no longitudinal Hall effect, then $R_{ikl} = -R_{kil}$ and it follows from (ii) that

$$R_1 = R_2 \qquad \text{and} \qquad R_4 = 0 \qquad \ldots\ldots\ldots \qquad \text{(iii)}$$

In the subsequent application to the bismuth crystal, we shall use a rectangular cartesian system of axes with OZ parallel to the principal (trigonal) axis and OX parallel to one of the binary axes. The main cleavage plane is thus the (x, y) plane. In the experiments which concern us, the magnetic field is invariably perpendicular to the current flow. For this case Shoenberg has shown that a longitudinal Hall effect, which is implicit in equation (i), produces a change in the specific resistance given by

$$\begin{aligned} \Delta\rho = H[&(R_2 - R_1) \cos\alpha \sin\alpha \sin\chi \\ &- R_4 \sin^2\alpha(\sin\chi \sin 3\zeta + \cos\chi \cos\alpha \cos 3\,\zeta)] \qquad \ldots \qquad \text{(iv)} \end{aligned}$$

Here α is the angle between the principal axis and the current direction, ζ is the angle between the binary axis and the projection of the current vector on the main cleavage plane, and χ is the angle between the magnetic field and the projection of the principal axis on the plane normal to the current vector. The significance of these angles can be seen from fig. 7 (*a*). Expression (iv) represents a magneto-resistive effect which is linear in H, and the change in the observed resistivity when the field is reversed should be $2\,\Delta\rho$. If the current flow is perpendicular or parallel to the main cleavage plane ($\alpha=0°$ or $90°$) then the effect vanishes, as is also the case in polycrystalline material.

In Shoenberg's own experiment $\alpha=45°$, $\zeta=17°$ and χ was varied by rotating the magnet. Reversal of the magnetic field produced no detectable change in resistance for any value of χ and the essentially transverse nature of the Hall effect was regarded as established. However, only one value of magnetic field (13,700 oersteds) was used and it is probable that this was too high for the effect to appear. The effect is exceedingly small and is clearly more easily detected in weak fields where the orthodox magnetic resistance is not large.

In the orientation discussed by Meixner (1939) the magnetic field is parallel to one of the binary axes and the current is restricted to the (y, z) plane. Thus $\chi=\zeta=90°$ and equation (iv) reduces to

$$\Delta\rho/\mathrm{H}=(\mathrm{R}_2-\mathrm{R}_1)\cos\alpha\sin\alpha+\mathrm{R}_4\sin^2\alpha \quad . \quad . \quad . \quad . \quad (\mathrm{v})$$

For $\alpha=45°$ (see fig. 7 (*b*)) this expression has a maximum value in agreement with Meixner. An identical arrangement, with $\alpha=45°$, was used by Casimir and Gerritsen (1941), and produced an apparent " Umkehreffekt"; however, such an effect was also obtained when the magnetic field was perpendicular to the binary axis, which disagrees with the theory.

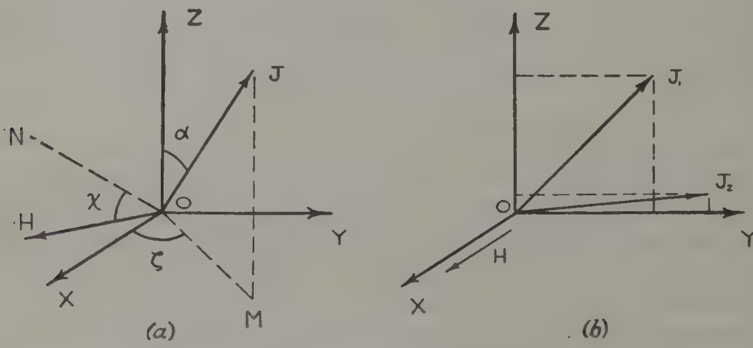
In discussing our own experimental results, the significance of the ϕ resistance curves must be borne in mind. Stierstadt (1933) has shown that in P_1 ($\perp$) crystals (current flow perpendicular to a binary axis) the resistance in moderately low fields is controlled by the orientation of the (111) plane, *i. e.* the main cleavage plane. When t is is parallel to the field the resistance has a maximum value, and when it is perpendicular to the field the resistance has a minimum value. Further, in P_1 (||) crystals, that is, current flow parallel to binary axis, the controlling influence on the resistance is exerted by the $(11\bar{1})$ plane, *i. e.* the subsidiary cleavage planes. The behaviour of these planes is opposite to that of the main cleavage plane. For example, if one of the subsidiary cleavage planes is parallel to the field the resistance has a minimum value. In this case the inclination of the main cleavage plane to the field is approximately 71° and therefore it is clear that in any P_1 type crystal the minimum of the ϕ curve represents a state of affairs in which the main cleavage plane is roughly perpendicular to the field; at any rate the departure must be less than 20°.

Returning to fig. 7 (*b*), we see that if the current vector **J** is parallel to the y-axis we have a normal P_1 ($\perp$) crystal with a main cleavage plane parallel to the field, *i. e.* at the maximum of the ϕ curve. If **J** is restricted to the (x, y) plane, any type of P_1 crystal can be represented and as $\alpha=90°$ in every case no "Umkehreffekt" is possible. As it has already been established in the previous paper that in these bismuth fibres the main cleavage plane is parallel to the axis and hence parallel to the current flow, we see why the majority of fibres do not exhibit this effect.

The explanation of the anomalous behaviour of the minority described in § 2 would seem to lie in the exceptional nature of the crystal structure of these fibres. In drawing the fibre the main cleavage plane has become tilted with respect to the normal orientation, with the result that the current flow is no longer parallel to the main cleavage plane and α is no longer 90°. This displacement must necessarily be quite small and of the order of a few degrees.

We consider first the orientation corresponding to the maximum of the ϕ curve. In P_1 ($\perp$) crystals the main cleavage plane is parallel to the field and, as shown in fig. 7 (*b*), the current vector is displaced slightly from the y-axis although still remaining in the (y, z) plane. For this case, therefore, equation (v) will apply. In P_1 ($\|$) crystals, and intermediate cases, both **H** and **J** will be inclined to the main cleavage plane and no corresponding simplification of equation (iv) can be made.

Fig. 7.



Orientation of current J and magnetic field H with respect to crystal axes, OZ parallel to principal axis; OX parallel to binary axis.

(*a*) General case; OM is projection of J on (x, y) plane; ON is projection of z-axis on plane normal to J.

(*b*) Special case with J in (y, z) plane and H parallel to OX. J_1 and J_2 correspond to values of 45° and approximately 85° respectively for α.

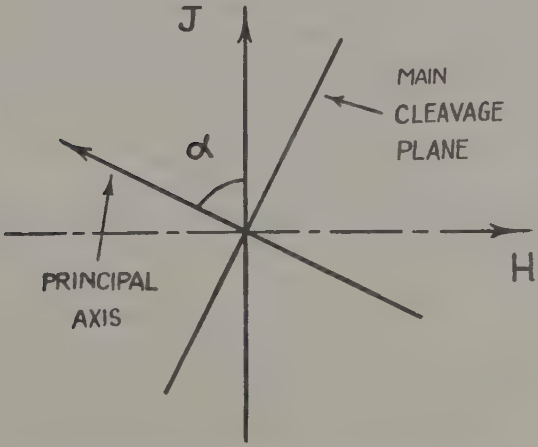
Next we consider the orientation corresponding to the minimum of the ϕ curve. In true P_1 ($\perp$) crystals the main cleavage plane would be perpendicular to the field; in imperfect P_1 ($\perp$) crystals, that is to say, crystals possessing a fault in accordance with the conjecture made above, this will no longer by the case, but the principal axis will be in the plane defined by **H** and **J**. It may be instructive to consider this orientation, which is shown in fig. 8 with the crystal fault intentionally exaggerated. If the principal axis lies in the plane of **H** and **J**, then $\chi=0$ and this restriction alone causes (iv) to reduce to

$$\Delta\rho/\mathrm{H}=-\mathrm{R}_4 \sin^2\alpha \cos\alpha \cos 3\zeta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad \text{(vi)}$$

The " Umkehreffekt " is controlled by α, *i.e.* the magnitude of the crystal fault, and by ζ, which is determined by the position of the binary axis. In crystals of the quasi-P_1 ($\perp$) type, the binary axis is normal to the current* and $\zeta=90°$. Hence the resistance change should be normal. For all other orientations of the binary axis expression (vi) has a finite value.

In crystals of the quasi-P_1 ($\|$) type the position is complicated by the fact that the principal axis is inclined at an angle of approximately 19° to the plane containing **H** and **J**. For intermediate P_1 crystals the angle will have some value between 0 and 19° and, since α, χ and ζ are all finite for these cases, equation (iv) must be used as it stands.

Fig. 8.



Crystal fault (exaggerated) in which main cleavage plane is not perpendicular to magnetic field. The principal axis, current J and field H all lie in one plane.

CONCLUSION.

The nature of the specimens we have used confers two very real advantages in exploring the reality of such an " Umkehreffekt." The preparation of crystals of such a relatively high resistance enables one to examine with care and precision any departures from the normal behaviour. This benefit is to some extent obtained at the expense of exactness in specifying to which crystalline category any given fibre belongs. Furthermore the size and the method of mounting would seem to eliminate the possibility that such observed effects were due to a component of transverse Hall coefficient.

---

* In this case fig. 8 represents a section in the (y, z) plane.

It is difficult to resist the conclusion that the evidence here presented confirms the existence in these specimens of a longitudinal component of the Hall coefficient. Because there is at the moment an inevitable lack of precision in our specification of the category of any given fibre, we can go no further than the above qualitative interpretation setting out the various possibilities which can give rise to an " Umkehreffekt." With reference to the order of magnitude, it is relevant to quote Casimir and Gerritsen (*loc. cit.*) : "A linear term would lead to a decrease of resistance at low fields but . . . . this decrease would only amount to $2\times10^{-4}\rho_0$." The results illustrated in figs. 2 and 3 are in striking agreement with this prediction.

REFERENCES.

Casimir, H. B. G. and Gerritsen, A. N., 1941, *Physica, 's Grav.*, **8**, 1108.
Conn, G. K. T. and Donovan, B., 1948, *Nature, Lond.*, **162**, 336.
Donovan, B. and Conn, G. K. T., 1949, *Phil. Mag.*, [7], **40**, 283.
Grüneisen, E. and Gielsessen, J., 1936, *Ann. Phys. Lpz.*, **26**, 449.
Kapitza, P., 1928, *Proc. Roy. Soc.* A, **119**, 358.
Kohler, M., 1934, *Ann. Phys. Lpz.*, **20**, 878, 891; 1941, *Ibid.*, **40**, 601.
Meixner, J., 1939, *Ann. Phys. Lpz.*, **36**, 105 ; 1941, *Ibid.*, **40**, 165.
Shoenberg, D., 1935, *Proc. Camb. Phil., Soc.*, **31**, 271.
Stierstadt, O., 1933 a, *Zeits. f. Physik*, **80**, 636 ; 1933 b, *Ibid.*, **85**, 310.
Thompson, N., 1938, *Proc. Roy. Soc.* A, **164**, 24.
Verleger, H., 1932, *Zeits. f. Physik*, **76**, 760.
Voigt. W., 1928, *Lehrbuch der Kristallphysik* (Leipzig : Teubner).

LXX. *The Experimental Determination of the Geomagnetic Radial Variation.*

By S. K. Runcorn, A. C. Benson and A. F. Moore,
The Dept. of Geodesy and Geophysics, Cambridge *
and
D. H. Griffiths,
University, Manchester †.

[Received June 8, 1950.]

Abstract.

Theories of the origin of the dipole components of the Earth's main magnetic field are of two types; distributed theories attribute it to a fundamental property of rotating matter, core theories to current systems within the core. The variation with depth below the surface of the horizontal field intensity is different for the two theories. Experimental values are given for this variation, obtained by measurements in coal mines, which are near to the values predicted by a core theory, and significantly different from those predicted by a distributed theory. The magnetic effects of the sedimentary rocks and magnetic anomalies due to the basement rocks are shown to have negligible results on the measurements.

It has been suggested, notably by Blackett (1947 and 1949), that the axial dipole component of the geomagnetic field might be a fundamental property of any rotating spherical mass; the non-dipole part of the field (about 15 per cent) arising from subsidiary processes such as induction. The other possible theory of the geomagnetic field suggests that it is due to the flow of electric currents maintained by thermo-electromotive forces or by dynamo action; these must be confined to the deep interior of the earth, where the electrical conductivity seems likely to be appreciably higher than that of the surface rocks. It is convenient to call the former a Distributed Theory and the latter a Core Theory.

Bullard suggested that it might be possible to decide between these two theories by measurements of the change with depth inside the earth of the horizontal and vertical components of the geomagnetic field. Preliminary measurements were made in mines in South Africa (Hales and Gough 1947) and in Lancashire (reported by Chapman 1948 a). Much more extensive measurements have now been made in the coal mines of this country and methods of establishing the significance of the results, which were lacking in the earlier work, have been developed.

---

* Formerly at the Physical Laboratories of the University of Manchester.

† Communicated by the Authors.

Were the distribution of the Earth's field exactly that of a dipole, the field components would vary radially according to an inverse cube law except in regions where the field vector is rotational. For a core theory such a region (of electric current flow) is inaccessibly deep within the Earth. The simplest interpretation of a distributed theory—though perhaps not the only one—is that the whole Earth is a region of virtual current flow, having magnetic but not electrical effects. Runcorn (1948) and Chapman (1948 b) calculated, for the two groups of theories, the differences (ΔZ, ΔH) between the vertical and horizontal field components at a small depth "d" below the Earth's surface (Z_d and H_d respectively) and the vertical and horizontal components at the surface (Z_0 and H_0 respectively). They showed that though for a Core Theory, if R is the radius of the Earth,

$$\Delta Z = Z_d - Z_0 = +\frac{3\cdot 0 \,.\, d}{R} \,.\, Z_0$$

and

$$\Delta H = H_d - H_0 = +\frac{3\cdot 0 \,.\, d}{R} \,.\, H_0,$$

a Distributed Theory, assuming the density of the surface rocks is 2·7, gives

$$\Delta Z = Z_d - Z_0 \doteq +\frac{3\cdot 0 \,.\, d}{R} \,.\, Z_0,$$

$$\Delta H = H_d - H_0 = -\frac{6\cdot 5 \,.\, d}{R} \,.\, H_0.$$

The essential problem of the experiment is to find conditions in which measured differences of the geomagnetic components between a point on the Earth's surface and a point underground may be attributed to the main field and not to anomalies. For though the geomagnetic field is mainly a dipole field, the gradients of its components over small distances at the surface may be very different from the dipole gradients. Thus in England, while the surface gradients to be expected from the dipole field are of the order of 5 gammas per kilometre (1 gamma$=10^{-5}$ gauss), the observed gradients in some regions may be of the order of hundreds of gammas per kilometre. These anomalous gradients seem to arise from surface features of the igneous basement rocks or from horizontal changes in their magnetite content; the sedimentary layers above usually producing little effect.

Therefore in regions where geological data indicates that the probable minimum depth at which extensive bodies of igneous rock are likely to occur is greater than 8000 feet, as in South Lancashire and Yorkshire, the gradients of the field components over distances of a few thousand feet, both along the Earth's surface and vertically, should be sensibly linear. If this is so, then the effect of anomalies arising from the basement rocks on the measured values of ΔZ and ΔH can be estimated by a surface survey and allowed for, if necessary.

For at any point where there is no current density (real or virtual) the gradients of the geomagnetic field along the 3 axes of a rectangular coordinate system are related by 4 equations, because

$$\text{curl } \mathbf{H}=0 \qquad \text{and} \qquad \text{div. } \mathbf{B}=0.$$

Such an equation is

$$\frac{d\text{H}}{dz}=\frac{d\text{Z}}{dx},$$

where z is a vertical distance and x a distance along a tangent to a line of longitude at the point.

The measurement of the surface gradients may prove difficult as the values of the field at points on the surface are likely to be perturbed by the field of igneous boulders in the drift clay, thus some scatter of the points from a straight line has to be expected. Also the measurements may be made impossible owing to the vicinity of built up areas. Thus the preliminary measurements at the Parsonage Colliery at Leigh in Lancashire are of less interest because no adequate surface survey could be made in the region.

It is difficult to find useful sites in mines as steel arches and rails are used in almost all passages and even if these are cleared some iron is usually buried. Passages also have to be a sufficient distance from the large concentrations of steel near the shafts, so that no gradient is detectable along the passage. Only then can the geomagnetic field in the passage be assumed to be free from man made disturbances. Only a very few of the cleared sites proved to be free of buried iron.

Measurements of the vertical and horizontal components are made to an accuracy of 1 gamma by the quasi-absolute B.M.Z. (magnetometric zero balance) and Q.H.M. (quartz horizontal magnetometer) respectively which were developed by Dr. La Cour at the Danish Meteorological Institute. The temperature coefficients of the instruments must be known accurately and frequently checked in the laboratory, because their values are of the order of 10 gammas per degree centigrade and the difference in temperature between the surface and underground sites may be as much as 30° centigrade. For a determination of ΔH and ΔZ measurements of H and Z are taken on the surface at a fixed point, the instruments are then taken to the underground site, the measurements repeated, and finally a further series is taken at the surface site. Throughout the experiment, instruments are left on the surface to record the temporal change in the geomagnetic field so that a correction for this may be applied to the results. The surface survey was made by H and Z variometers of the Schmidt balance type over a grid of 4·5 kms. around the surface site and from these measurements the surface gradients were computed and the value of the field at the surface site was corrected for any local anomaly.

The results of the measurements are shown in Tables I.–V. Table I. shows the position of the underground and surface sites where measurements were made, the depth of the former and the height above Ordnance Datum of the latter. The distances of these sites from the shafts and the steel materials and buildings associated with them are also given. Only in the case of the Nook surface site is there any interference, due to railway-sidings, and the correction for this is determined from the surface survey of the area. This Nook surface site was also used as the surface site for the Astley Green measurements, the necessary correction again being found from the surface survey of the area.

TABLE I.

Colliery	Latitude and longitude and National grid coordinates of underground site	National grid coordinates of surface site	Depth of underground site (feet)	Height above O.D. of surface (site) (feet)	Distance from shafts of underground site (yards)	Distance from shafts of surface site (yards)
Hickleton Main (near Doncaster, Yorks.).	53° 32′ 42″ N. 1° 17′ 36″ W. 44684056	44704065	2520	210	515	1200
Nook (near Leigh, Lancs.).	53° 29′ 50″ N. 2° 28′ 0″ W. 36924004	36914009	2930	100	560	410
Astley Green (near Leigh, Lancs.).	53° 28′ 40″ N. 2° 26′ 48″ W. 37053980	36914009	2859	64	1900	410

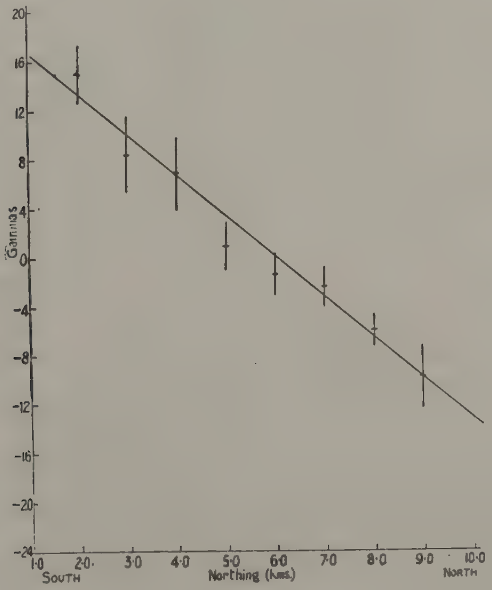
Table II. shows the results of the surface survey of the Lancashire and Yorkshire Areas. Values of H and Z on a kilometer grid covering a square with sides 4·5 km. from the surface base were used to compute, by a least squares method, the field gradients along each grid line. The gradients along the lines of latitude and longitude were then averaged. Figs. 1–4 show plots of H and Z against distances along lines of longitude (x) and along lines of latitude (y). The drawn straight line and the values given on the graph were derived by the method of least squares. The survey in South Lancashire was less satisfactory than that in Yorkshire, as small hamlets prevented measurements being made near certain of the grid points and the East–West gradient is rather large. However the corrections involved in ΔH and ΔZ are small.

TABLE II.

Surface Gradients of Geomagnetic Elements.

Colliery	Theoretical value of gradient of Z γ/km.	Experimental value of gradient of Z γ/km.	Theoretical value of gradient of H γ/km.	Experimental value of surface gradient of H γ/km.
	Variation with Latitude (Positive to North).			
Hickleton	$+5{\cdot}0$	$+5{\cdot}0 \pm 0{\cdot}3$	$-4{\cdot}0$	$-3{\cdot}4 \pm 0{\cdot}5$
Nook and Astley Green	$+5{\cdot}0$	$+4{\cdot}0$	$-4{\cdot}0$	$-6{\cdot}0$
	Variation with Longitude (Positive to East).			
Hickleton	0	$+1{\cdot}6 \pm 0{\cdot}5$	0	$+1{\cdot}2 \pm 0{\cdot}6$
Nook and Astley Green	0	$+3{\cdot}2$	0	$-8{\cdot}0$

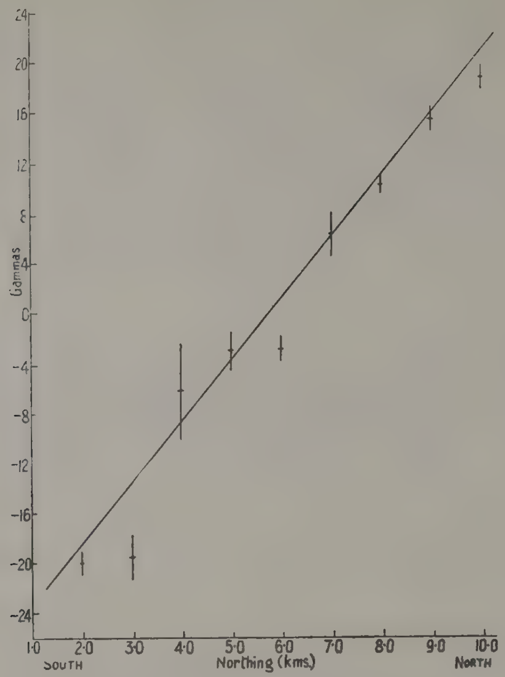
Fig. 1.



Hickleton Surface Survey.

Variation of H with latitude $\frac{\partial H}{\partial x} = -3{\cdot}4 \pm 0{\cdot}5\gamma | \text{km}.$

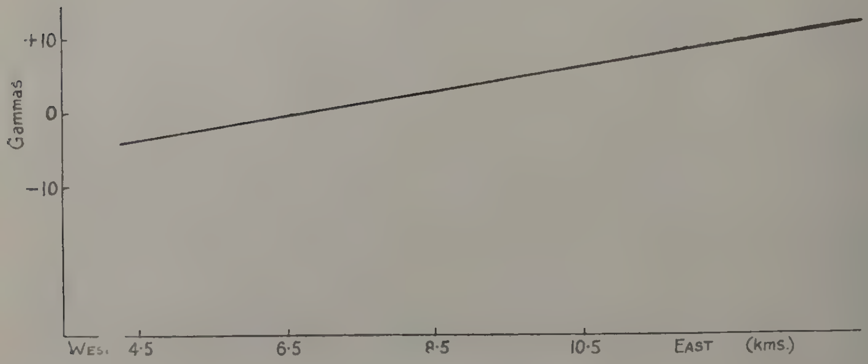
Fig. 2.



Hickleton Surface Survey.

Variation of z with latitude $\frac{\partial Z}{\partial x} = +5{\cdot}0 \pm 0{\cdot}3\gamma |$ km.

Fig. 3.

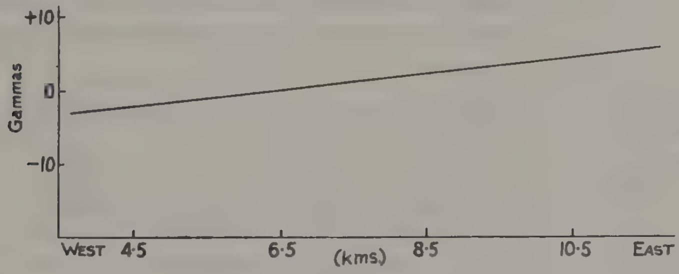


Hickleton Surface Survey.

Variation of z with longitude $\frac{\partial Z}{\partial y} = +1{\cdot}6 \pm 0{\cdot}5\gamma |$ km.

The effects on the measurements of the induced magnetism of the sedimentary rocks was examined. From boreholes and shafts in Lancashire and Yorkshire representative collections of strata samples were obtained and the magnetic susceptibilities of the specimens were determined with an astatic magnetometer. Table III. gives the average susceptibilities for the different types of rock.

Fig. 4.



Hickleton Surface Survey.

Variation of H with longitude $\frac{\partial H}{\partial y} = +1{\cdot}2 \pm 0{\cdot}6\gamma \,|\, \mathrm{km}$.

TABLE III.

(K is volume magnetic susceptibility.)

Locality	Mudstones, shales and siltstones		Sandstones		Fireclays		Ironstones	
	Number of samples	Average $K.10^6$	Number of samples	Average $K.10^6$	Number of samples	Average $K.10^6$	Number of samples	Average $K.10^6$
Lancashire	40	23·5	21	14·5	7	19·4	14	56·7
Yorkshire	60	23·4	11	11·7	6	14·1	4	69·7

Calculations were made of the magnetic effects of the strata at Hickleton, Assuming that the various beds have uniform susceptibility and thickness. the vertical field components due to each bed at the surface and the underground observation points could be determined, if the horizontal extent of the bed were known. The effects could then be summed for the whole succession, the difference in the underground and surface values being a measure of the anomaly in ΔZ. The horizontal extent of each bed is not known, but if it is assumed to be that for which its effect is a maximum, we get an anomaly which cannot possibly be exceeded but which is certain to be very greatly overestimated. A more realistic estimate is obtained by computing an average value of the anomalies which would be produced by beds of all sizes up to a certain arbitrary maximum. The "maximum" and "average" anomalies in ΔZ are shown in Table IV. The susceptibilities of the strata from the surface down to 2700 ft. have been measured and the susceptibilities of the strata beneath these down to the base of the Coal Measures at 4000 ft. have been estimated. The effect of all the sediments below the Coal Measures is negligible.

TABLE IV.

Thickness of Strata over which the effect is summed (feet from surface)	Vertical Component of Magnetic Fields due to Strata (gammas)					
	Maximum Effect			Average Effect		
	At surface	Underground (2367 feet)	Difference	At surface	Underground (2367 feet)	Difference
0–2728	+16	+30	+14	+5	+7	+2
2728–4000	0	+4	+4	0	+4	+4
Total			+18			+6

TABLE V.

(Positive Values denote Increase in Component with Depth).

	Theoretical ΔZ Both theories γ	Experimental ΔZ γ	Theoretical ΔH Core γ	Theoretical ΔH Distributed γ	Experimental ΔH γ
Hickleton	+16	+11±3	+6	−13	+22±2
Nook	+19	+11±3	+8	−16	+ 7±2
Astley Green	+18	+18±3	+7	−15	+10±3

Table V. shows the expected change with depth of the horizontal and vertical components on both the distributed and core theories together with the final experimental values. The measured values have been corrected so that they represent the difference between the components in the mine and at a point immediately above on the surface. The values of the geomagnetic components to be used at the surface are deduced from the survey and those underground from the survey of the passage in which the site is situated. It will be seen that the measuremets are near to those predicted for a Core Theory and do not support a Distributed Theory. Only in the case of the ΔH values at Hickleton do the results deviate significantly (*i. e.* by twice the standard deviation) from the prediction of a Core Theory.

Twenty sites in various mines have been magnetically surveyed but only six have been found to be free from buried iron. The result of the experiments in three of these sites has been described above. The experiment is proceeding in two more mines in the Doncaster area in order to obtain more ΔH measurements and a site in a colliery near Sunderland, Co. Durham is also being used. The detailed experimental results will be published elsewhere.

The experiments were done mainly from the Physics Dept. of Manchester University but are being continued from the Dept. of Geodesy and Geophysics of the University of Cambridge. We wish to thank Professor P. M. S. Blackett for his support and interest and Dr. Bruckshaw, the Danish Meteorological Institute, the Air Ministry, the Royal Greenwich Observatory, Dr. Veldkamp and Professor Whetton for the loan of instruments at various stages in the work. We are also grateful to the National Coal Board and its officials for providing facilities for the experiments and to many undergraduates of Manchester University for their assistance in the measurements.

REFERENCES.

BLACKETT, P. M. S., 1947, *Nature, Lond.*, **159**, 658 ; 1949, *Phil. Mag.* [7], **40**, 125.

CHAPMAN, S., 1948 a, *Nature, Lond.*, **161**, 462 ; 1948 b, *Annales de Geophysique*, **4**, 109.

HALES, A. L., and GOUGH, D. I., 1947, *Nature, Lond.*, **160**, 746.

RUNCORN, S. K., 1948, *Proc. Phys. Soc.*, **61**, 373.

LXXI. *Energies and Widths of Domain Boundaries in Ferromagnetics.*

By B. A. LILLEY,
Physics Department, University of Leeds*.

[Received May 30, 1950.]

ABSTRACT.

A systematic treatment is given of the energies and widths of domain boundaries in ferromagnetics; it covers cubic crystals, with positive or negative magneto-crystalline anisotropy coefficients, and crystals with a single easy direction of magnetization. This extends the investigations of Bloch (1932), Landau and Lifshitz (1935), Lifshitz (1944) and Néel (1944) on cobalt and iron, and enables estimates to be made of the energies and widths of domain boundaries in nickel. The effect of magneto-elastic anisotropy energy is considered. Theoretical values of the energies and widths of the more important types of boundary in nickel, iron and cobalt are calculated.

§1. INTRODUCTION.

A FERROMAGNETIC material is, in general, divided into elementary domains in each of which the spontaneous magnetization vector, $\mathbf{I}_0$, is constant, except possibly for small fluctuations in direction or intensity which may be neglected in the present calculations. The magnitude of $\mathbf{I}_0$ is the same in all the domains and is equal, to a good approximation, to the saturation intensity of the material at the temperature considered; the direction of $\mathbf{I}_0$ varies from domain to domain. The transition region between contiguous domains is known as a boundary or wall and in it $\mathbf{I}_0$ rotates from its direction in one domain to that in the other.

Estimates of the energy associated with domain boundaries and of their effective thickness were first made by Bloch in 1932. Since then the treatment has been refined and further developed by Landau and Lifshitz (1935), Néel (1944) and Lifshitz (1944). Landau and Lifshitz considered a crystal, such as hexagonal cobalt, with a single easy direction of magnetization; Néel and Lifshitz considered body-centred cubic iron, for which the easy directions lie along the cube edges, corresponding to a positive magneto-crystalline anisotropy coefficient.

The aim of this paper is to obtain more general formulae from which results may be obtained for all the particular cases of interest, and to apply similar methods to the more difficult problem of nickel, which has not previously been considered. For nickel, which has a face-centred

---

* Communicated by Professor E. C. Stoner, F.R.S.

structure, the easy directions lie along the cube diagonals, corresponding to a negative magneto-crystalline anisotropy coefficient. Most ferromagnetic materials are either cubic, with positive or negative anisotropy coefficients, or hexagonal, and these are all covered by the general formulæ obtained for the relative energies and widths of different types of boundary. Estimates of absolute values, such as are made for nickel, iron and cobalt, and could be made similarly for materials for which the basic experimental data are available, are much more uncertain.

In §2 the method of approach is indicated ; §3 deals with the energies and §4 with the widths of the domain boundaries neglecting the effect of magnetostriction. This is considered in §5, and in §6 the theoretical results for nickel, iron and cobalt are collected together.

A survey of the background to the subject of this paper is given by Stoner (1948, 1950 ; see in particular 1950, §3). Reference may also be made to a review article by Kittel (1949).

§2. METHOD.

In an unstrained ferromagnetic material in zero applied field each domain is magnetized along one of the crystallographically and energetically equivalent easiest directions. In domains in cobalt the magnetization vector, $\mathbf{I}_0$, can only lie along the hexagonal axis and thus only 180° boundaries (*i. e.* boundaries in which $\mathbf{I}_0$ rotates through 180°) can occur. In iron the domains can be magnetized along any one of the six equivalent cube edges, of type [100], and 90° and 180° boundaries are possible. Domains in nickel can be magnetized along a cube diagonal, of type [111], giving rise to 70·53°, 109·47° and 180° boundaries.

Néel (1944) emphasizes the importance of the condition that I_n, the normal component of $\mathbf{I}_0$, should have the same value not only on each side of the boundary, but also in it. This corresponds to the absence of free poles (*i. e.* a zero value of div **I**) in the transition region. The presence of internal free poles would in general correspond to a free energy very much greater than that for some other arrangement satisfying the condition of zero divergence. It is convenient to specify the direction of magnetization at any point in the boundary by (θ, ϕ), where θ is the angle between $\mathbf{I}_0$ and the normal to the boundary, and ϕ the angle between the projection of $\mathbf{I}_0$ on the boundary plane and some selected zero direction in this plane. In particular the initial and final directions of $\mathbf{I}_0$ in the adjacent domains may be denoted by (θ, ϕ_1) and (θ, ϕ_2), constancy of I_n being equivalent to constancy of θ.

The free energy per unit volume of the boundary may be expressed in the form

$$\mathrm{F}=\mathrm{F}_a+\mathrm{F}_e, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

where F_a is anisotropy and F_e exchange energy. The field energy is zero since the applied field and div **I** are both zero (*cf.* Stoner 1950, equations (3.14), (3.15)). The energy and effective width of the domain boundary are determined from the condition that F, integrated over the range in which ϕ changes from ϕ_1 to ϕ_2, shall be a minimum. The anisotropy energy, F_a, may be expressed in the form

$$F_a = F_c + F_m, \quad \ldots \ldots \quad (2.2)$$

or

$$\beta f_a = \beta f_c + \beta' f_m, \quad \ldots \ldots \quad (2.3)$$

where F_c is magneto-crystalline and F_m magneto-elastic anisotropy energy; f_a, f_c and f_m are expressions for reduced energies adjusted, by subtracting appropriate constant terms, to make each zero for $\mathbf{I}_0$ along (θ, ϕ_1) and (θ, ϕ_2); β, β' stand for the appropriate magneto-crystalline and magneto-elastic anisotropy energy density coefficients respectively. The effect of the term F_m will not be considered until § 5.

General expressions for boundary energy and width.

It is convenient to consider the boundary to be in the yz plane with its normal along Ox and to take Oy as the zero direction for measurement of ϕ. If $\delta\psi$ is the angle between the spins in adjacent layers $\delta\psi = \sin\theta\,\delta\phi$, and the exchange energy, which is proportional to $1-\cos\delta\psi$, may be written (Néel 1944, p. 7; *cf.* Stoner 1950, equation (3.1))

$$F_e = \alpha a^2 \sin^2\theta (d\phi/dx)^2, \quad \ldots \ldots \quad (2.4)$$

where α is proportional to the exchange energy for parallel spins, and a is a lattice constant. With the Weiss treatment of ferromagnetism α would be proportional to $\frac{1}{2}N_w I_0^2$, N_w being the molecular field coefficient.

The energy, γ, per unit area of the boundary is then given by

$$\gamma = \int \{\alpha a^2 \sin^2\theta (d\phi/dx)^2 + \beta f_a\}\, dx. \quad \ldots \ldots \quad (2.5)$$

It follows from a simple variational treatment that, for γ to have a minimum value,

$$\beta f_a = \alpha a^2 \sin^2\theta (d\phi/dx)^2, \quad \ldots \ldots \quad (2.6)$$

i.e. the exchange and anisotropy contributions to the energy are equal (see Néel 1944, pp. 6–8; *cf.* Becker and Döring 1939, p. 191; Stoner 1950, equations (3.1)–(3.6)). Equation (2.6) may be written

$$(dx/d\phi) = b_0 \sin\theta f_a^{-\frac{1}{2}}, \quad \ldots \ldots \quad (2.7)$$

giving, by substitution in (2.5),

$$\gamma/\gamma_0 = 2 \sin\theta \int_{\phi_1}^{\phi_2} f_a^{\frac{1}{2}}\, d\phi, \quad \ldots \ldots \quad (2.8)$$

where b_0 is a unit of length normal to the boundary, and γ_0 a unit for the energy per unit area of the boundary (*cf.* Stoner 1950, equation (3.7)) given by

$$b_0 = (\alpha a^2/\beta)^{\frac{1}{2}}; \quad \gamma_0 = (\alpha a^2 \beta)^{\frac{1}{2}}. \quad \ldots \ldots \quad (2.9)$$

Equations (2.8) and (2.7) will be used to calculate domain boundary energies and widths.

Magneto-crystalline anisotropy energy.

For a cubic crystal the magneto-crystalline energy, F_c, apart from constant terms, may be represented, with sufficient approximation, by (*cf.* Stoner 1950, equation (2.1))

$$F_c=K_4(\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2+\alpha_1^2\alpha_2^2). \qquad (2.10)$$

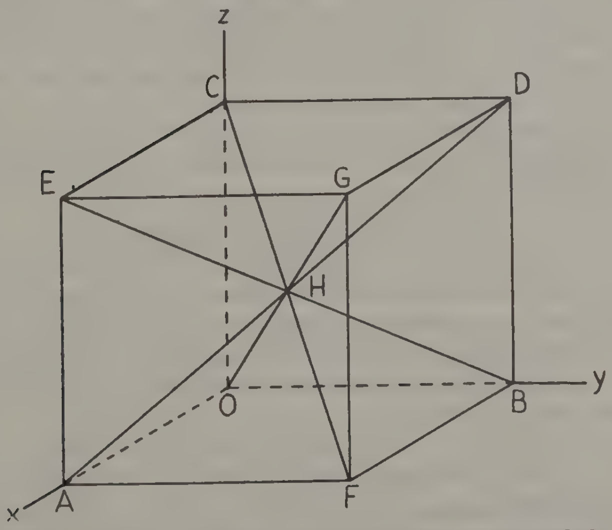
where α_1, α_2, α_3 are the direction cosines of $\mathbf{I}_0$ relative to the cube edges of the crystal. The reduced energy f_c is a function of θ and ϕ and may be written

$$\beta f_c(\theta, \phi)=(F_c)_{\theta,\phi}-(F_c)_{\min}, \qquad (2.11)$$

where $(F_c)_{\theta,\phi}$ is the value of F_c corresponding to $\mathbf{I}_0$ along the direction (θ, ϕ); $(F_c)_{\min}$ that corresponding to $\mathbf{I}_0$ along an easy direction; and for cubic crystals, $\beta=|K_4|$. For iron K_4 is positive and thus

$$(F_c)_{\min}=0. \qquad (2.12)$$

Fig. 2.1.



Cubic crystal, showing easy directions in iron (edges) and in nickel (diagonals).

For nickel K_4 is negative giving

$$(F_c)_{\min}=K_4/3=-\beta/3. \qquad (2.13)$$

The form of the function f_c can be determined for boundary planes of all possible orientations, and in the most general case a very complicated expression is obtained, but only certain boundaries satisfy the condition imposed upon I_n. All 90° boundaries in iron are similar to those in which $\mathbf{I}_0$ rotates from [100] (*i. e.* OA in fig. 2.1) to [010] (OB); and $[11l]$ or $[\bar{1}\bar{1}l]$ (*i. e.* any direction in the plane OCGF) are the only directions which make equal angles with these cube edges. In nickel all 109·47°

boundaries are similar to those in which $\mathbf{I}_0$ rotates from $[\bar{1}11]$ (HD) to $[1\bar{1}1]$ (HE) and, again, only $[11l]$ or $[\bar{1}\bar{1}l]$ make equal angles with these cube diagonals. All 70·53° boundaries in nickel are similar to those in which $\mathbf{I}_0$ rotates from $[1\bar{1}1]$ (HE) to $[1\bar{1}\bar{1}]$ (HA); and $[hk0]$ (*i. e.* any direction in the plane parallel to OAFB drawn through H) are the only directions making equal angles with these initial and final easy directions. All 180° boundaries in iron are similar to those formed by a rotation of $\mathbf{I}_0$ from [001] to $[00\bar{1}]$ and in nickel from [111] to $[\bar{1}\bar{1}\bar{1}]$, the normals to the boundaries thus lying in the (001) and (111) planes respectively. Thus, for 90° boundaries in iron possible boundary planes are parallel to the $\{11l\}$ planes ($\{11l\}$ being understood, in the usual crystallographic notation, to mean any plane of the type $(11l)$); for 109·47° boundaries in nickel to the $\{11l\}$ planes; and for 70·53° boundaries in nickel to the $\{hk0\}$ planes. It is convenient to consider {001} to be $\{11l\}$ planes with $l=\infty$; the parameters h, k and l can each take any value between $-\infty$ and $+\infty$. The {001}, {110} and {111} planes are particular examples of the $\{11l\}$ planes; {100} and {110} of either the $\{11l\}$ or $\{hk0\}$ planes. In these cases f_c assumes particularly simple forms and it is then possible to derive analytical expressions for the boundary energies and widths. The normals to all 180° boundaries in iron and nickel lie in the {001} and {111} planes respectively.

For hexagonal cobalt, the magneto-crystalline energy may be represented, with sufficient approximation, by

$$\mathrm{F}_c=\beta f_c(\theta,\phi)=\mathrm{K}_2'\sin^2\phi, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.14)$$

where $\beta=\mathrm{K}_2'$. The angle θ is equal to 90° and hence ϕ is the angle between $\mathbf{I}_0$ and the hexagonal axis. This relation also holds, if the natural magneto-crystalline effect can be neglected, for materials of positive or negative isotropic magnetostriction under tension or compression respectively.

Magneto-crystalline anisotropy energy of cubic crystals as a function of θ and ϕ.

If the normal to the boundary is the $[11l]$ direction it is possible to determine the direction cosines of $\mathbf{I}_0$ with respect to the axes $[11l]$, $[\bar{1}10]$, $[\bar{l}l2]$. A simple transformation then gives the direction cosines with respect to the [100], [010], [001] axes defined by the cube edges, and it follows from (2.10) that

$$\begin{aligned}\mathrm{F}_c=\frac{\mathrm{K}_4}{4(l^2+2)^2}[&4(2l^2+1)+4(l^2+1)(l^2-4)\sin^2\theta-(3l^2+4)(l^2-4)\sin^4\theta\\ &-4(l^2-1)\{6+(l^2-4)\sin^2\theta\}\sin^2\theta\sin^2\phi\\ &+4(l^2+3)(l^2-1)\sin^4\theta\sin^4\phi\\ &-2^{5/2}\,l\cos\theta\sin\theta\sin\phi\{2(l^2-1)\\ &-(5l^2+4)\sin^2\theta+2(l^2+5)\sin^2\theta\sin^2\phi\}]. \quad . \quad . \quad . \quad (2.15)\end{aligned}$$

This relation holds for $\{11l\}$ boundary planes in any crystal whose magneto-crystalline energy may be represented by (2.10). In particular, $l=\infty$, 0, 1 correspond to boundary planes whose normals are [001], [110], [111] respectively. The energy expression is now in a convenient form for use in calculating boundary energies and widths.

§ 3. BOUNDARY ENERGIES.

The energy of a domain boundary when magneto-crystalline anisotropy and exchange energy alone are considered, given by replacing f_a by f_c in (2.8), is

$$\gamma/\gamma_0=2\sin\theta\int_{\phi_1}^{\phi_2} f_c^{\frac{1}{2}}\,d\phi. \qquad \ldots\ldots\ldots \quad (3.1)$$

Nickel. For a boundary plane in nickel whose normal is $[11l]$ the values of θ and ϕ_3, ϕ_4 corresponding to $\mathbf{I}_0$ along the $[\bar{1}11]$, $[1\bar{1}1]$ directions respectively, can be determined in terms of l:

$$\theta=\cos^{-1} l/\{3(l^2+2)\}^{\frac{1}{2}}, \qquad 0<\theta<\pi\,; \ldots\ldots\ldots \quad (3.2)$$

$$\left.\begin{aligned}&\phi_3=\cos^{-1}\{(l^2+2)/(l^2+3)\}^{\frac{1}{2}}, \quad 0<\phi_3<\pi/2\,;\\ &\phi_4=\pi-\phi_3.\end{aligned}\right\} \quad \ldots \quad (3.3)$$

The domain boundaries in nickel which are of particular importance are now considered, and the results are summarized in Table 3.1. A [001], 70·53° boundary is understood to mean a 70·53° boundary the normal to whose plane lies in the direction [001], about which $\mathbf{I}_0$ rotates from the initial to the final easy direction. The values of θ, ϕ_3 and ϕ_4 are obtained by substituting the appropriate value of l in (3.2) and (3.3), ϕ_1 and ϕ_2 are found from ϕ_3 and ϕ_4, and $f_c^{\frac{1}{2}}$ is then determined by substitution in (2.15) and using (2.11) and (2.13). The relative energies, (γ/γ_0), of the boundaries are then calculated using (3.1).

The [001], 109·47° boundary is effectively two successive [001], 70·53° boundaries since $\mathbf{I}_0$ must rotate through $[\bar{1}\bar{1}1]$ (*i. e.* HC in fig. 2.1) in passing from $[\bar{1}11]$ (HD) to $[1\bar{1}1]$ (HE). Its energy is, therefore, twice that of the [001], 70·53° boundaries. It is impossible to have [111], 70·53° boundaries since the angles between [111] and any two easy directions containing an angle of 70·53° (*e. g.* HD and HC in fig. 2.1) are not equal.

180° *Boundaries.* All 180° boundaries are similar to those formed by a rotation of $\mathbf{I}_0$ from [111] to $[\bar{1}\bar{1}\bar{1}]$. A new system of axes $Ox'y'z'$ which has directions [111], $[\bar{1}10]$, $[\bar{1}\bar{1}2]$ (corresponding to $l=1$ in the $[11l]$, $[\bar{1}10]$, $[\bar{l}\bar{l}2]$ axes of § 2) with respect to the system, $Oxyz$, formed by the cube edges of the crystal, may be set up. A particular boundary may be specified by ω if $\mathbf{I}_0$ rotates through OP, which lies in the $y'z'$ plane, and where ω is the angle between OP and Oy' (see fig. 3.1). The boundary thus formed is in the plane whose normal is $[3^{\frac12}\sin\omega-\cos\omega, -3^{\frac12}\sin\omega-\cos\omega, 2\cos\omega]$. The angle ω can have any value and, in particular, when $\omega+30°=n60°$, where n is any integer, $\mathbf{I}_0$ rotates through an easy direction intermediate between [111] and $[\bar{1}1\bar{1}]$. It is thus convenient to write *

$$\lambda=\omega+30°, \qquad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (3.4)$$

and the normal to the boundary plane as $[\sin\lambda-3^{\frac12}\cos\lambda, -2\sin\lambda, \sin\lambda+3^{\frac12}\cos\lambda]$. The angle λ is that between the projections of $\mathbf{I}_0$ and the easy direction $[\bar{1}1\bar{1}]$ on the (111) plane.

TABLE 3.1.

Reduced energies of 70·53° and 109·47° domain boundaries in nickel.
θ, angle between $\mathbf{I}_0$ and Ox, the normal to the boundary.
ϕ_1, ϕ_2, angles between projection of $\mathbf{I}_0$ on yz plane and Oy, corresponding to initial and final easy directions respectively.
f_c, reduced magneto-crystalline energy.
γ, boundary energy per unit area.
$\gamma_0=(\alpha a^2\beta)^{\frac12}$ (see (2.9)).
$\eta=\cos^{-1}(2/3)^{\frac12}(=35{\cdot}26°)$: 2η=angle between [111] and $[\bar{1}11]$.

Type of boundary	Easy directions; initial, final.	θ	ϕ_1, ϕ_2.	$3^{5/2}2f_c^{\frac12}$	$3^{\frac12}(\gamma/\gamma_0)$	(γ/γ_0)
[001], 70·53°	$[\bar{1}11]$, $[\bar{1}\bar{1}1]$.	$\frac12\pi-\eta$	0, $\pi/2$.	$2^2 3^{3/2}\sin\phi\cos\phi$	$2^{3/2}/3$	0·5443
[001], 109·47°	$[\bar{1}11]$, $[1\bar{1}1]$.	$\frac12\pi-\eta$	0, π.	$2^2 3^{3/2}\sin\phi\cos\phi$	$2^{5/2}/3$	1·0887
[110], 70·53°	$[\bar{1}1\bar{1}]$, $[\bar{1}11]$.	$\frac12\pi$	$-\eta$, η.	$3^2(1-3\sin^2\phi)$	$2^{\frac12}-\eta$	0·4611
[110], 109·47°	$[\bar{1}11]$, $[1\bar{1}1]$.	$\frac12\pi$	η, $\pi-\eta$.	$3^2(3\sin^2\phi-1)$	$2^{\frac12}+\pi/2-\eta$	1·3680
[111], 109·47°	$[\bar{1}11]$, $[1\bar{1}1]$.	2η	$\pi/6$, $5\pi/6$.	$2^{7/2}(1-\sin 3\phi)^{\frac12}$	$2^{15/2}/3^4$	1·2903

The magneto-crystalline energy may be represented by

$$F_c=(K_4/12)\{4\cos^4\phi+3\sin^4\phi+2^{5/2}\cos\phi\sin^3\phi\sin\omega(3–4\sin^2\omega)\}, \quad (3.5)$$

---

* This angle, following the usage of Néel (1944), is denoted by λ. It is not to be confused with a magnetostriction coefficient, which is always written in this paper with a directional suffix, *e. g.*, λ_{100}.

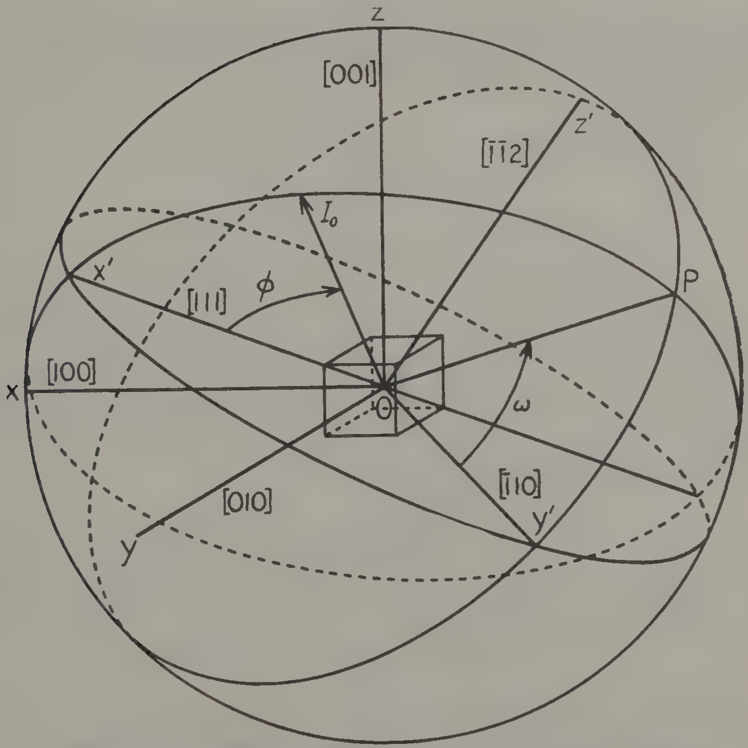
by putting $l=1$ and replacing θ by ϕ and ϕ by ω in (2.15). The magnetization vector, $\mathbf{I}_0$, must rotate in the plane of the boundary and $\sin\theta$ in (3.1) is, therefore, equal to unity. From (2.11), (2.13), (3.1) and (3.5) it follows that

$$\gamma/\gamma_0=3^{-\frac{1}{2}}\int_0^{\pi}(1+\rho\sin\phi\cos\phi+7\cos^2\phi)^{\frac{1}{2}}\sin\phi\,d\phi, \quad . \quad . \quad (3.6)$$

where

$$\left.\begin{aligned}\rho&=-2^{5/2}\sin 3\omega\\ &=2^{5/2}\cos 3\lambda\end{aligned}\right\};\ -2^{5/2}\leqslant\rho\leqslant+2^{5/2}. \quad . \quad . \quad . \quad (3.7)$$

Fig. 3.1.



Coordinate axes used in describing 180° boundaries in nickel.
OP, direction in $y'z'$ plane through which $\mathbf{I}_0$ rotates.
ω, angle between OP and Oy'.
ϕ, angle between $\mathbf{I}_0$ and Ox'.

It is possible, without any difficulty, to evaluate the integral analytically when $\rho=\pm 2^{5/2}$ since $(\sin^2\phi\pm 2^{5/2}\sin\phi\cos\phi+8\cos^2\phi)$ is a perfect square. Similarly when $\rho=0$ the integral can be evaluated analytically in terms of an inverse trigonometric hyperbolic function. For a general value of ρ the integration is slightly more troublesome.

It is possible to make the mixed term of (3.6) equal to zero by writing

$$\phi=\chi+\kappa, \qquad (3.8)$$

where χ is a new variable and κ is a function of ρ only, and choosing κ such that

$$\tan 2\kappa=\rho/7. \qquad (3.9)$$

Then

$$\gamma/\gamma_0=3^{-\frac{1}{2}}(7\sec 2\kappa)^{\frac{1}{2}}\left\{\int_{-\kappa}^{\pi-\kappa}(r^2-\sin^2\chi)^{\frac{1}{2}}\sin\kappa\cos\chi\,d\chi\right.$$
$$\left.+\int_{-\kappa}^{\pi-\kappa}(r^2-\sin^2\chi)^{\frac{1}{2}}\cos\kappa\sin\chi\,d\chi\right\}. \qquad (3.10)$$

The first integral is evaluated by substituting $u=(\sin\chi)/r$ and the second by substituting $v=(\cos\chi)/s$ and, after some manipulation, the energy of a boundary whose normal is $[\sin\lambda-3^{\frac{1}{2}}\cos\lambda,\ -2\sin\lambda,\ \sin\lambda+3^{\frac{1}{2}}\cos\lambda]$ may be written

$$\gamma/\gamma_0=3^{-\frac{1}{2}}\{2^{3/2}+(7\sec 2\kappa)^{\frac{1}{2}}(r^3t\sin^{-1}t+s^3p\sinh^{-1}p)\}, \qquad (3.11)$$

where

$$\left.\begin{aligned}14r^2&=9\cos 2\kappa+7\\ 14s^2&=9\cos 2\kappa-7\end{aligned}\right\};\ s^2=r^2-1;\ \left.\begin{aligned}t&=(\sin\kappa)/r\\ p&=(\cos\kappa)/s.\end{aligned}\right\} \qquad (3.12)$$

When $\lambda=0°, 120°, 240°$, $\rho=2^{5/2}$; $\lambda=60°, 180°, 300°$, $\rho=-2^{5/2}$, and these six particular cases correspond to boundary planes of the (110) type and to a rotation of $\mathbf{I}_0$ through an easy direction intermediate between [111] and $[\bar{1}\bar{1}\bar{1}]$. Thus the energy of these boundaries is equal to the sum of the energies of [110], 70·53° and [110], 109·47° boundaries (*cf.* Table 3.1) and by substitution in (3.11)

$$\gamma/\gamma_0=3^{-\frac{1}{2}}(2^{3/2}+\pi/2-2\eta)=1{\cdot}8292. \qquad (3.13)$$

This is the minimum value for the energy of 180° boundaries in nickel. When $\lambda=30°$, $\rho=0$, the normal to the boundary is $[\bar{1}\bar{1}2]$ and the energy as given by (3.11) is a maximum,

$$\gamma/\gamma_0=3^{-\frac{1}{2}}\{2^{3/2}+7^{-\frac{1}{2}}\sinh^{-1}(7^{\frac{1}{2}})\}=2{\cdot}0040. \qquad (3.14)$$

Iron. For a boundary plane whose normal is $[11l]$ the values of θ and ϕ_3, ϕ_4 corresponding to $\mathbf{I}_0$ along the [100], [010] directions are given by

$$\theta=\cos^{-1}(l^2+2)^{-\frac{1}{2}}\qquad ,\ 0<\theta<\pi; \qquad (3.15)$$

$$\left.\begin{aligned}\phi_3&=\cos^{-1}-\{(l^2+2)/2(l^2+1)\}^{\frac{1}{2}},\ \pi<\phi_3<3\pi/2;\\ \phi_4&=\cos^{-1}\{(l^2+2)/2(l^2+1)\}^{\frac{1}{2}}\ \ ,\ 3\pi/2<\phi_4<2\pi.\end{aligned}\right\} \qquad (3.16)$$

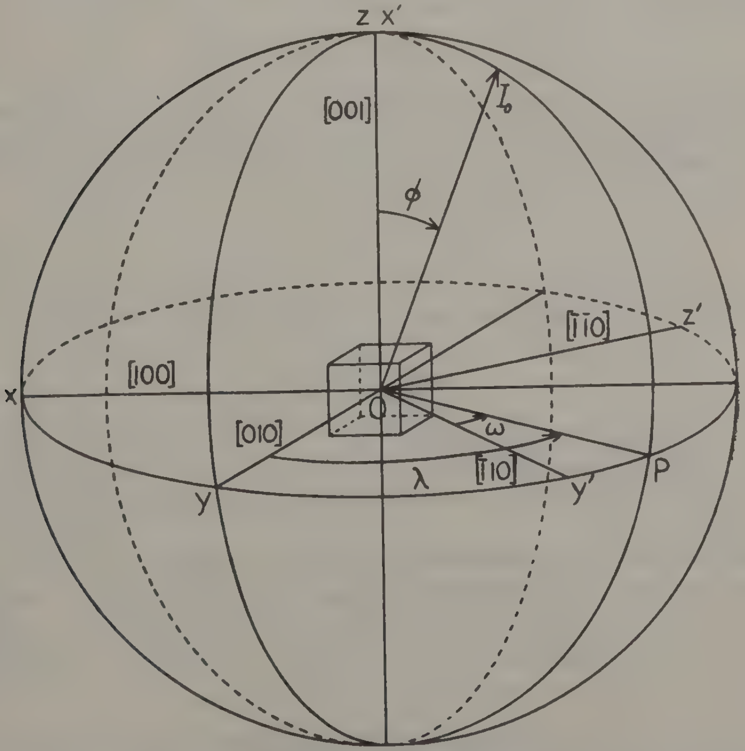
The domain boundary energies may be calculated in a way similar to that used for nickel, using (3.15), (3.16) and (2.12) instead of (3.2), (3.3) and (2.13). The results are summarized in Table 3.2.

TABLE 3.2.

Reduced energies of 90° domain boundaries in iron.
In all cases the initial and final easy directions are [100] and [010] respectively. (See Table 3.1 for symbols.)

Type of boundary	θ	$\left.\begin{matrix}\phi_1,\\ \phi_2.\end{matrix}\right\}$	$4f_c^{\frac{1}{2}}$	γ/γ_0	γ/γ_0
[001], 90°	$\pi/2$	$\left.\begin{matrix}-3\pi/4,\\ -\pi/4.\end{matrix}\right\}$	$2(1-2\sin^2\phi)$	1	1·0000
[110], 90°	$\pi/4$	$\left.\begin{matrix}-\pi,\\ 0.\end{matrix}\right\}$	$\sin\phi(8-3\sin^2\phi)^{\frac{1}{2}}$	$1+2^{-3/2}3^{-\frac{1}{2}}5\sinh^{-1}(3/5)^{\frac{1}{2}}$	1·7274
[111], 90°	$\pi/2-\eta$	$\left.\begin{matrix}-5\pi/6,\\ -\pi/6.\end{matrix}\right\}$	$2^3 3^{-3/2}(1+\sin 3\phi)^{\frac{1}{2}}$	$2^5/3^3$	1·1852

Fig. 3.2.



Coordinate axes used in describing 180° boundaries in iron.
OP, direction in xy ($y'z'$) plane through which $\mathbf{I}_0$ rotates.
ω, angle between OP and Oy'.
λ, angle between OP and Oy.
ϕ, angle between $\mathbf{I}_0$ and Oz (Ox').

180° *Boundaries.* A rotation of $\mathbf{I}_0$ from [001] to [00$\bar{1}$] through OP, which is in the xy plane (see fig. 3.2) is considered by substituting $l=\infty$ and replacing θ by ϕ and ϕ by ω in (2.15). By writing

$$\lambda=\omega+45° \qquad (3.17)$$

(*cf.* equation (3.4)) the magneto-crystalline anisotropy energy may be expressed by

$$f_c=\sin^2\phi-\mu\sin^4\phi, \qquad (3.18)$$

where

$$\left.\begin{aligned}\mu&=(7-\cos 4\omega)/8,\\ &=(7+\cos 4\lambda)/8 \quad (\textit{cf.}\ \text{Néel 1944, p. 11}).\end{aligned}\right\} \qquad (3.19)$$

The angle λ is that between [010] and the projection of $\mathbf{I}_0$ on the (001) plane.

The normal to the boundary specified by λ is $[-\cos\lambda, -\sin\lambda, 0]$. By substitution in (3.1) it follows that

$$\gamma/\gamma_0=2\{1+(1-\mu)\mu^{-\frac{1}{2}}\sinh^{-1}(\mu/1-\mu)^{\frac{1}{2}}\}. \qquad (3.20)$$

When $\lambda=n\pi/2$, where n is any integer, $\mu=1$, and the boundary can be considered to be the sum of two [001], 90° type boundaries. Its energy as calculated from (3.20) is twice that of a [001], 90° boundary and is given by

$$\gamma/\gamma_0=2. \qquad (3.21)$$

This is the minimum value for the energy of 180° boundaries in iron. If $\lambda=n\pi/2+45°$, corresponding to $\mu=3/4$, the energy is a maximum and equal to

$$\gamma/\gamma_0=2+3^{-\frac{1}{2}}\sinh^{-1}(3^{\frac{1}{2}})=2{\cdot}7603 \qquad (3.22)$$

for boundaries in the {110} planes.

The relative energies of domain boundaries in iron have already been calculated by Néel (1944), whose method is essentially the same as that employed above.

Cobalt. Only 180° boundaries occur in hexagonal cobalt and F_c is zero along the easy directions. Thus, from (2.14),

$$f_c=\sin^2\phi, \qquad (3.23)$$

giving, from (3.1), writing $\sin\theta=1$,

$$\begin{aligned}\gamma/\gamma_0&=2\int_0^{\pi}\sin\phi\,d\phi\\ &=4. \qquad (3.24)\end{aligned}$$

The energies of all boundaries in cobalt are the same.

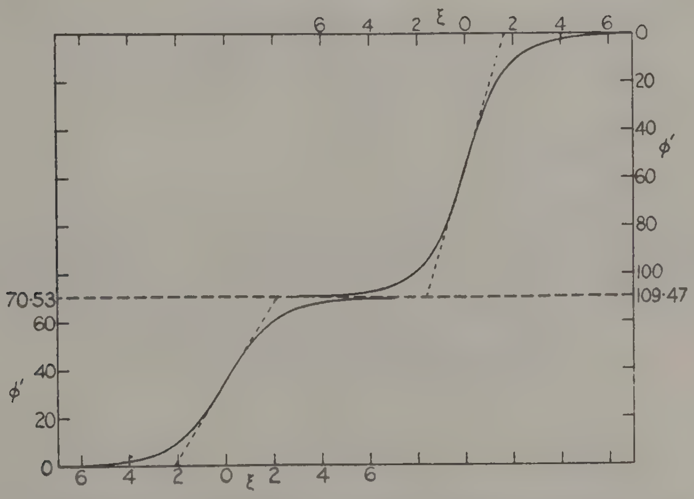
§ 4. Boundary Widths.

If exchange and magneto-crystalline anisotropy energy alone are considered the distance x, measured normal to the boundary, can be obtained as a function of ϕ, the angle between the projection of $\mathbf{I}_0$ on the boundary plane and the zero direction, by replacing f_a by f_c in equation (2.7), giving

$$(dx/d\phi)=b_0 \sin\theta \, f_c^{-\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

This section will be devoted to a consideration of the variation of x with ϕ for the various boundaries considered in § 3, and of the resultant "boundary widths".

Fig. 4.1.



Variation with distance normal to the boundary plane of the angle, ϕ', through which $\mathbf{I}_0$ has rotated, for [110], 70·53° and [110], 109·47° domain boundaries in nickel.

The broken line is an asymptote common to the two curves. $\xi=x/b_0$. x, distance normal to boundary.

For $\phi\rightarrow\phi_1$ or ϕ_2, $f_c\rightarrow 0$, $|x|\rightarrow\infty$, and if the boundary is to satisfy the minimum energy condition it has no sharply defined width, as is shown in figs. 4.1, 4.2 and 4.3. It will be convenient to introduce a non-dimensional parameter (*cf.* Stoner 1950, equation (3.8)),

$$\xi=x/b_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.2)$$

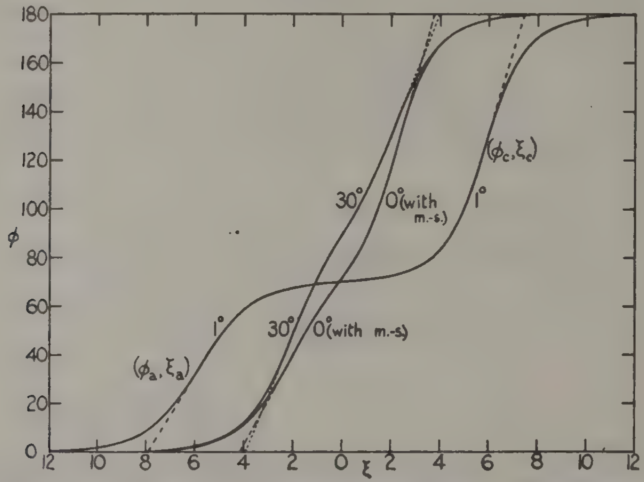
for measurement of length normal to the boundary. All ϕ, ξ curves for 70·53° and 109·47° nickel, 90° iron and 180° cobalt boundaries are symmetrical about their points of inflexion which lie at $\phi=(\phi_1+\phi_2)/2$. For such boundaries the effective width, b, may be conveniently taken as

$$(b/b_0)=(\phi_2-\phi_1)(d\xi/d\phi)_{\text{inflexion}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.3)$$

where the notation is self-explanatory. All ϕ, ξ curves for 180° boundaries in nickel and iron are of the form indicated in figs. 4.2 and 4.3 respectively, and the effective width, b, may be taken as

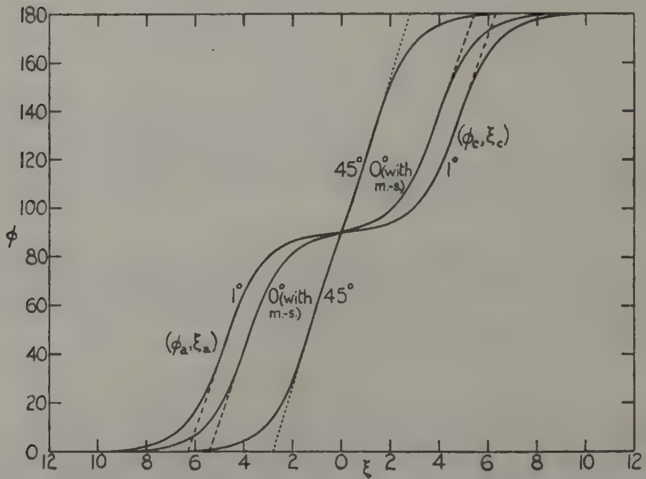
$$(b/b_0)=(\xi_c-\xi_a)+\phi_a(d\xi/d\phi)_a+(\pi-\phi_c)(d\xi/d\phi)_c, \quad . \quad . \quad . \quad (4.4)$$

Fig. 4.2.



Variation with ξ of ϕ for 180° boundaries in nickel. The numbers on the curves indicate the values of λ; the curve marked 0° (with m.-s.) is that for $\lambda=0°$ when magnetostriction is taken into account (see § 5).
$\xi=x/b_0$. x, distance normal to boundary.
ϕ, angle between $\mathbf{I}_0$ and [111].

Fig. 4.3.



Variation with ξ of ϕ for 180° boundaries in iron. The numbers on the curves indicate the values of λ; the curve marked 0° (with m.-s.) is that for $\lambda=0°$ when magnetostriction is taken into account (see § 5).
$\xi=x/b_0$. x, distance normal to boundary. ϕ, angle between $\mathbf{I}_0$ and [001].

where (ϕ_a, ξ_a), (ϕ_c, ξ_c) (see figs. 4.2 and 4.3) are the first and third points of inflexion. It will be noted that all 180° boundaries in iron, and 180° boundaries in nickel when $\lambda = n60° + 30°$ (where λ is defined as the angle between the projections of $\mathbf{I}_0$ and the easy direction $[\bar{1}1\bar{1}]$ on the (111) plane (see (3.4)), are symmetrical about $\phi = \pi/2$. The expressions (4.1) and (4.2) are used to find ξ as a function of ϕ, the constant of integration being omitted, and (4.3), (4.4) are used to determine the boundary widths. The results for 70·53° and 109·47° boundaries in nickel are summarized in Table 4.1 and those for 90° boundaries in iron in Table 4.2. The values of θ, ϕ_1 and ϕ_2 and $f_c^{\frac{1}{2}}$ are given in Tables 3.1 and 3.2 and are not repeated here.

TABLE 4.1.

Reduced widths of 70·53° and 109·47° domain boundaries in nickel.
$\xi = x/b_0$.
x, distance normal to boundary.
b, effective boundary width.
b_0, unit for measurement of x and b; $b_0 = (\alpha a^2/\beta)^{\frac{1}{2}}$ (see (2.9)).
$\eta = \cos^{-1}(2/3)^{\frac{1}{2}} (= 35{\cdot}26°)$.

Type of boundary	$(2/3)^{\frac{1}{2}}\xi$	$(2/3)^{\frac{1}{2}}(b/b_0)$	(b/b_0)
[001], 70·53°	$ln\,\lvert\tan\phi\rvert$	π	3·8476
[001], 109·47°		∞	∞
[110], 70·53°	$2\tanh^{-1}(2^{\frac{1}{2}}\tan\phi)$	$2^{5/2}\eta$	4·2642
[110], 109·47°	$-2\coth^{-1}(2^{\frac{1}{2}}\tan\phi)$	$2^{\frac{1}{2}}(\pi-2\eta)$	3·3093
[111], 109·47°	$ln\,\lvert\tan(\pi/8-3\phi/4)\rvert$	π	3·8476

TABLE 4.2.

Reduced widths of 90° domain boundaries in iron.
(See Table 4.1 for symbols.)

Type of boundary	ξ	(b/b_0)	(b/b_0)
[001], 90°	$ln\,\lvert\tan(\phi+\pi/4)\rvert$	π	3·1416
[110], 90°	$-\sinh^{-1}\{(8/5)^{\frac{1}{2}}\cot\phi\}$	$(8/5)^{\frac{1}{2}}\pi$	3·9738
[111], 90°	$-ln\,\lvert\tan(3\pi/8-3\phi/4)\rvert$	π	3·1416

It may be noted, in connection with the expressions for ξ in the Tables, that the inverse hyperbolic functions may be expressed as logarithmic functions as follows:

$$\left.\begin{aligned} \sinh^{-1}x &= ln\{x+(x^2+1)^{\frac{1}{2}}\}\,; \\ 2\tanh^{-1}x &= ln\{(1+x)/(1-x)\},\ x^2<1\,; \\ 2\coth^{-1}x &= ln\{(x+1)/(x-1)\},\ x^2>1\,; \\ 2\tanh^{-1}(\tan\phi) &= ln\,\lvert\tan(\phi+\pi/4)\rvert\,. \end{aligned}\right\} \quad . \quad . \quad (4.5)$$

Values of these functions are tabulated by Comrie (1949).

If the magnetostriction term, f_m, is not introduced, the [001], 109·47° boundary in nickel can only be regarded as two successive 70·53° boundaries, and as these would necessarily have an infinite distance between them, the resultant boundary width would be infinite. This case will be considered in § 5.

In fig. 4.1 the ϕ, ξ curves for [110], 70·53° and [110], 109·47° boundaries in nickel have been plotted. All ϕ, ξ curves for 70·53° and 109·47° nickel and 90° iron boundaries are similar to these.

180° *Boundaries in nickel.* The parameter ξ is expressed as a function of ϕ by using (4.1), (4.2), (3.5), (2.11) and (2.13) and may be written

$$\xi=3^{\frac{1}{2}}2\int(1+\rho\cot\phi+8\cot^2\phi)^{-\frac{1}{2}}\operatorname{cosec}^2\phi d\phi \quad . \quad . \quad . \quad . \quad (4.6)$$

$$\left.\begin{aligned}&=-(3/2)^{\frac{1}{2}}\sinh^{-1}\{(16\cot\phi+\rho)/(32-\rho^2)^{\frac{1}{2}}\}\\&=-(3/2)^{\frac{1}{2}}\sinh^{-1}\{\operatorname{cosec}3\lambda(2^{3/2}\cot\phi+\cos 3\lambda)\}.\end{aligned}\right\} \quad . \quad . \quad (4.7)$$

The ϕ, ξ curves corresponding to $\lambda=1°$ and 30°, where λ is the angle between the projections of $\mathbf{I}_0$ and the easy direction $[\bar{1}1\bar{1}]$ on the (111) plane, are plotted in fig. 4.2. The curve for $\lambda=30°$ is a limiting case; and as λ approaches 0° the central portion broadens and becomes more nearly parallel to the ξ axis until, when $\lambda=0°$, the boundary width is infinite and the wall can be regarded as two infinitely distant walls of the form shown in fig. 4.1. When

$$\left.\begin{aligned}&\lambda=\ 0°,(b/b_0)=\infty\,;\\&\lambda=\ 1°,(b/b_0)=15{\cdot}4029\,;\\&\lambda=30°,(b/b_0)=2(3/2)^{\frac{1}{2}}\{\sinh^{-1}(6^{\frac{1}{2}})+(7/2)^{\frac{1}{2}}\sin^{-1}(2/7^{\frac{1}{2}})\}\\&\qquad\qquad=7{\cdot}9161.\end{aligned}\right\} \quad . \quad (4.8)$$

180° *Boundaries in iron.* For iron ξ may be written, by using (4.1), (4.2) and (3.18), in the form

$$\xi=\int(\sin^2\phi-\mu\sin^4\phi)^{-\frac{1}{2}}d\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.9)$$

$$=-\sinh^{-1}\{(\cot\phi)/(1-\mu)^{\frac{1}{2}}\}=-\sinh^{-1}(2\operatorname{cosec}2\lambda\cot\phi). \quad (4.10)$$

Curves, similar to those for 180° nickel boundaries, are plotted in fig. 4.3, that for $\lambda=45°$ (λ being the angle between the projection of $\mathbf{I}_0$ on the (001) plane and the easy direction [010]), corresponding to a normal in the direction $[\bar{1}\bar{1}0]$, being a limiting case. In the other limiting case $\lambda=0°$ and the boundary can be considered to be two infinitely distant boundaries of the [001], 90° type. When

$$\left.\begin{aligned}&\lambda=0°,\qquad (b/b_0)=\infty\,;\\&\lambda=1°,\qquad (b/b_0)=12{\cdot}6258\,;\\&\lambda=22{\cdot}5°,\quad (b/b_0)=2\{\sinh^{-1}(6^{\frac{1}{2}})+(7/2)^{\frac{1}{2}}\sin^{-1}(2/7^{\frac{1}{2}})\}\\&\qquad\qquad\qquad=6{\cdot}4635\,;\\&\lambda=45°,\qquad (b/b_0)=2\{\sinh^{-1}(2^{\frac{1}{2}})+3^{\frac{1}{2}}\sin^{-1}(2/3)^{\frac{1}{2}}\}\\&\qquad\qquad\qquad=5{\cdot}6018.\end{aligned}\right\} \quad . \quad (4.11)$$

Boundaries in cobalt. For cobalt ξ may be written, by using (4.1), (4.2) and (3.23), in the form

$$\xi=\int \operatorname{cosec}\phi\, d\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.12)$$

$$=ln\,|\tan(\phi/2)|\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.13)$$

This curve is similar to those for 70·53° and 109·47° nickel and 90° iron boundaries. The reduced boundary width, b/b_0, is equal to π and is the same for all the boundaries.

§5. Effect of Magnetostriction.

The introduction of a magneto-elastic anisotropy energy term, F_m, has important consequences. This section will be devoted to a consideration of its effect on 180° boundaries, and particularly to those which have been considered in §§3 and 4 to be the sums of two infinitely distant boundaries. They no longer have infinite " boundary widths " since the effective result of the introduction of this term is an attraction between neighbouring 109·47° and 70·53° nickel boundaries and between adjacent 90° boundaries in iron (see Stoner 1950, § 3 (i) ; Kittel 1949, pp. 563–4).

The strain tensor and the energy term $\beta' f_m$. The magneto-elastic energy, F_m, of a cubic crystal is the sum of magnetostrictive and elastic parts and may be written (Becker and Döring 1939, p. 146)

$$\left.\begin{aligned} F_m=\{&-3C_2\lambda_{100}(\alpha_1^2A_{11}+\alpha_2^2A_{22}+\alpha_3^2A_{33})\\ &-3C_3\lambda_{111}2(\alpha_1\alpha_2A_{12}+\alpha_2\alpha_3A_{23}+\alpha_3\alpha_1A_{31})\}\\ &+\{\tfrac{1}{2}C_1(A_{11}+A_{22}+A_{33})^2+C_2(A_{11}^2+A_{22}^2+A_{33}^2)\\ &+2C_3(A_{12}^2+A_{23}^2+A_{31}^2)\}, \end{aligned}\right\} \quad . \quad (5.1)$$

where the A_{ik} are components of the strain tensor, and λ_{100}, λ_{111} are magnetostriction coefficients. Comparison of (5.1) with Bitter (1937, p. 172, equation (138)) shows that the constants C_1, C_2, C_3 are given in terms of the elastic coefficients c_{11}, c_{12}, c_{44} by

$$C_1=c_{12}\,; \qquad C_2=(c_{11}-c_{12})/2\,; \qquad C_3=c_{44}. \quad . \quad . \quad . \quad (5.2)$$

The stable configuration of the crystal may be determined by minimizing F_m with respect to the A_{ik}. It follows that

$$A_{ik}=(3/2)\lambda_{111}\alpha_i\alpha_k, \qquad A_{ii}=(3/2)\lambda_{100}\{\alpha_i^2-C_1/(3C_1+2C_2)\}, \quad . \quad (5.3)$$
$$i\neq k;$$

from which it is evident that the components, A_{ik}, of the strain tensor are the same for domains magnetized in opposite directions. It may be shown further (*cf.* Lifshitz 1944, pp. 338–9) that they remain constant throughout the 180° boundary, *i. e.* the deformation of the material of the boundary is effectively the same as that in the adjacent domains. The A_{ik} are found by substitution in (5.3) of the direction cosines corresponding to the easy directions in the domains.

The excess magneto-elastic energy of the boundary, $\beta' f_m$, with respect to the domains can now be calculated by substitution in (5.1) of the appropriate values of $[\alpha_1\alpha_2\alpha_3]$, corresponding to the directions of $\mathbf{I}_0$ in the boundary and in the domains. (See Lifshitz 1944, p. 339; Kittel 1949, p. 564.)

Nickel. The directions of magnetization in adjacent domains are [111] and $[\bar{1}\bar{1}\bar{1}]$ (*cf.* §§ 3 and 4) and the strain tensor has components

$$A_{12}=A_{23}=A_{31}=\lambda_{111}/2\,; \qquad A_{11}=A_{22}=A_{33}=C_2\lambda_{100}/(3C_1+2C_2). \quad (5.4)$$

For a rotation of $\mathbf{I}_0$ through OP (see fig. 3.1) the values of α_1, α_2 and α_3 in the boundary are

$$\left.\begin{aligned} \alpha_1&=6^{-\frac{1}{2}}\{2^{\frac{1}{2}}\cos\phi-(3^{\frac{1}{2}}\sin\lambda+\cos\lambda)\sin\phi\},\\ \alpha_2&=6^{-\frac{1}{2}}\{2^{\frac{1}{2}}\cos\phi+2\cos\lambda\sin\phi\},\\ a_3&=6^{-\frac{1}{2}}\{2^{\frac{1}{2}}\cos\phi+(3^{\frac{1}{2}}\sin\lambda-\cos\lambda)\sin\phi\}, \end{aligned}\right\} \quad . \quad . \quad . \quad (5.5)$$

giving, for the excess magneto-elastic energy, an expression,

$$\beta' f_m=(9/2)C_3\lambda_{111}^2\sin^2\phi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.6)$$

which is independent of the angle λ. The reduced anisotropy energy may be written, from (2.3), (2.11), (2.13), (3.5), (3.7) and (5.6), in the form

$$f_a=\{(1+\tau)/12\}\{1+\rho'\sin\phi\cos\phi+\sigma\cos^2\phi\}\sin^2\phi, \quad . \quad . (5.7)$$

where

$$\tau=(54C_3\lambda_{111}^2)/|K_4|\,; \qquad \rho'=\rho/(1+\tau)\,; \qquad \sigma=7/(1+\tau). \quad . \quad . \quad (5.8)$$

Substitution in (2.8), with $\sin\theta$ equal to unity, and comparison with (3.6) gives

$$\left.\begin{aligned} \gamma/\gamma_0&=\{(1+\tau)/3\}^{\frac{1}{2}}\int_0^{\pi}(1+\rho'\sin\phi\cos\phi+\sigma\cos^2\phi)^{\frac{1}{2}}\sin\phi\,d\phi,\\ &=\{(1+\tau)/3\}^{\frac{1}{2}}\{(\sigma+1)^{\frac{1}{2}}+(\sigma\sec 2\kappa)^{\frac{1}{2}}(r_1^3t_1\sin^{-1}t_1\\ &\qquad\qquad +s_1^3p_1\sinh^{-1}p_1)\}, \end{aligned}\right\} \quad . \quad . \quad (5.9)$$

where

$$\left.\begin{aligned} 2\sigma r_1^2&=(\sigma+2)\cos 2\kappa+\sigma\\ 2\sigma s_1^2&=(\sigma+2)\cos 2\kappa-\sigma \end{aligned}\right\}\,; \qquad s_1^2=r_1^2-1\,; \qquad \left.\begin{aligned} t_1&=(\sin\kappa)/r_1\\ p_1&=(\cos\kappa)/s_1, \end{aligned}\right\} \quad (5.10)$$

and $\tan 2\kappa=\rho'/\sigma=\rho/7$.

The elastic coefficients of single nickel crystals have been measured by Bozorth *et al.* (1949). Values of the magnetostriction coefficients are given by Becker and Döring (1939, pp. 277–80), and of the magnetocrystalline anisotropy coefficient, K_4, by Stoner (1950, § 2). The constants required are

$$\left.\begin{array}{ll} c_{11}=2{\cdot}50\times10^{12}\text{ erg cm.}^{-3}, & \lambda_{100}=-46\times10^{-6}, \\ c_{12}=1{\cdot}60\times10^{12}\text{ erg cm.}^{-3}, & \lambda_{111}=-25\times10^{-6}, \\ c_{44}=1{\cdot}185\times10^{12}\text{ erg cm.}^{-3}, & K_4=-0{\cdot}50\times10^{5}\text{ erg cm.}^{-3}, \end{array}\right\} \quad (5.11)$$

and the value of τ calculated from them is 0·80.

When

$$\rho=\pm2^{5/2},\quad \gamma/\gamma_0=3^{-1/2}[(8+\tau)^{1/2}+\{(9+\tau)/9\}\sin^{-1}(9+\tau)^{-1/2} +(2^{3/2}\tau/9)\sinh^{-1}(8/\tau)^{1/2}],$$

$$=2{\cdot}19\,; \quad (5.12)$$

$$\rho=0,\quad \gamma/\gamma_0=\{(1+\tau)/3\}^{1/2}\{(\sigma+1)^{1/2}+\sigma^{-1/2}\sinh^{-1}(\sigma^{1/2})\},$$

$$=2{\cdot}27. \quad (5.13)$$

These energies are greater than those calculated in §3 where the effect of magnetostriction is neglected.

The parameter ξ may now be expressed as a function of ϕ, using (5.7), by

$$\xi=-2\{3/(8+\tau)\}^{1/2}\sinh^{-1}[\{2(8+\tau)\cot\phi+\rho\}/\{4(8+\tau)(1+\tau)-\rho^2\}^{1/2}]. \quad (5.14)$$

When

$$\left.\begin{array}{ll} \rho=\pm2^{5/2}, & b/b_0=7{\cdot}91\,; \\ \rho=0, & b/b_0=4{\cdot}45. \end{array}\right\} \quad (5.15)$$

The boundary width no longer approaches infinity as ρ approaches $\pm2^{5/2}$ and the introduction of a magneto-elastic energy term produces effectively an attraction between [110], 70·53° and [110], 109·47° boundaries.

Iron. The directions of $\mathbf{I}_0$ in adjacent domains are [001] and [00$\bar{1}$] (*cf.* §§ 3 and 4) and the strain tensor has components (*cf.* Lifshitz 1944, pp. 338–9 ; Kittel 1949, p. 564)

$$A_{12}=A_{23}=A_{31}=0\,;\ A_{11}=A_{22}=-3\lambda_{100}C_1/2(3C_1+2C_2)\,;$$

$$A_{33}=3\lambda_{100}(C_1+C_2)/(3C_1+2C_2). \quad (5.16)$$

For a rotation through OP (see fig. 3.2) the direction of $\mathbf{I}_0$ in the boundary is $[-\sin\lambda\sin\phi, \cos\lambda\sin\phi, \cos\phi]$ and, hence,

$$\beta' f_m=(9/2)C_2\lambda_{100}{}^2\sin^2\phi, \quad (5.17)$$

which is independent of the angle λ. The reduced anisotropy energy may be written in the form

$$f_a = (1+\tau')(\sin^2\phi - \mu'\sin^4\phi), \quad . \quad . \quad . \quad . \quad . \quad (5.18)$$

where

$$\tau' = (9C_2\lambda_{100}^2)/2K_4\,; \quad \mu' = \mu/(1+\tau'). \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.19)$$

It follows that

$$\gamma/\gamma_0 = 2(1+\tau')^{\frac{1}{2}}[1+(1-\mu')\mu'^{-\frac{1}{2}}\sinh^{-1}\{\mu'/(1-\mu')\}^{\frac{1}{2}}]; \quad . \quad (5.20)$$

and that

$$\left.\begin{aligned} \xi &= -(1+\tau')^{-\frac{1}{2}}\sinh^{-1}\{(1-\mu')^{-\frac{1}{2}}\cot\phi\}\,; \\ b/b_0 &= 2(1+\tau')^{-\frac{1}{2}}[\sinh^{-1}\{(2\mu'-1)/(1-\mu')\}^{\frac{1}{2}} + 2\mu'^{\frac{1}{2}}\sin^{-1}(2\mu')^{-\frac{1}{2}}]. \end{aligned}\right\} \quad (5.21)$$

The constants c_{11}, c_{12}, c_{44} are given by Kittel (1949, p. 556), λ_{100}, λ_{111} by Becker and Döring (1939, p. 280), and K_4 by Stoner (1950, § 2) as follows :

$$\left.\begin{aligned} c_{11} &= 2{\cdot}41\times10^{12}\,\mathrm{erg\ cm^{-3}}, & \lambda_{100} &= 19{\cdot}5\times10^{-6}, \\ c_{12} &= 1{\cdot}46\times10^{12}\,\mathrm{erg\ cm^{-3}}, & \lambda_{111} &= -18{\cdot}8\times10^{-6}, \\ c_{44} &= 1{\cdot}12\times10^{12}\,\mathrm{erg\ cm^{-3}}, & K_4 &= 4{\cdot}7\times10^{5}\,\mathrm{erg\ cm^{-3}}. \end{aligned}\right\} \quad . \quad (5.22)$$

The value of C_2 obtained from these is $0{\cdot}48\times10^{12}$ erg cm^{-3} giving a value of

$$\tau' = 0{\cdot}0017.$$

When

$$\left.\begin{aligned} \lambda &= 0, \\ \mu &= 1, \end{aligned}\right\} \gamma/\gamma_0 = 2\{(1+\tau')^{\frac{1}{2}} + \tau'\sinh^{-1}(\tau'^{-\frac{1}{2}})\}$$

$$= 2{\cdot}02, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.23)$$

$$\begin{aligned} b/b_0 = {} & 2(1+\tau')^{-\frac{1}{2}}[\sinh^{-1}\{(1-\tau')/\tau'\}^{\frac{1}{2}} \\ & + 2(1+\tau')^{-\frac{1}{2}}\sin^{-1}\{(1+\tau')/2\}^{\frac{1}{2}}] \end{aligned}$$

$$= 10{\cdot}87\,; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.24)$$

$$\left.\begin{aligned} \lambda &= 45°, \\ \mu &= 3/4, \end{aligned}\right\} \gamma/\gamma_0 = 2[(1+\tau')^{\frac{1}{2}} + \{(1+4\tau')/3^{\frac{1}{2}}2\}\sinh^{-1}\{3/(1+4\tau')\}^{\frac{1}{2}}]$$

$$= 2{\cdot}77, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.25)$$

$$\begin{aligned} b/b_0 = {} & 2(1+\tau')^{-\frac{1}{2}}[\sinh^{-1}\{(2-4\tau')/(1+4\tau')\}^{\frac{1}{2}} \\ & + \{3^{\frac{1}{2}}/(1+\tau')\}\sin^{-1}\{2(1+\tau')/3\}] \end{aligned}$$

$$= 5{\cdot}59. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.26)$$

Thus the inclusion of a term representing the magneto-elastic energy of the crystal has little effect on the value of the domain boundary energy (*e.g.* 2·02 as compared with 2·00 for a [001], 180° boundary; for other values see Table 6.1), but results, as in the case of nickel, in an important modification of the results obtained, in § 4, for the boundary widths.

In this section the numerical values of τ and τ' are given to two significant figures; those of γ/γ_0 and b/b_0 are given to three figures. The accuracy of these values is, however, determined by the accuracy with which the constants concerned (given in (5.11), (5.22)) have been measured.

Cobalt. The effect of magnetostriction in cobalt is not important since only 180° boundaries are possible and none of these has an infinite width even if magneto-elastic energy is neglected. Moreover, the introduction of this further term would have very little effect on the boundary energy since the magnetostriction coefficients of cobalt are smaller than those for iron, and the magneto-crystalline anisotropy coefficient, β, is much greater (by a factor of about 10) than that for iron.

§ 6. COLLECTED RESULTS AND CONCLUSION.

In conclusion, it is appropriate to tabulate the results obtained. This is done in Tables 6.1 and 6.2, in which " with m.-s." is to be understood to mean " with magnetostriction taken into account ". In cobalt the hexagonal axis lies in the domain boundaries.

TABLE 6.1.

Reduced energies, γ/γ_0, of domain boundaries in ferromagnetics.
γ, boundary energy per unit area.
For estimated values of γ_0 see text.

Type of boundary		Direction of normal to boundary			
		[001]	[110]	[111]	$[\bar{1}\bar{1}2]$
Ni	70·53°	0·5443	0·4611	—	—
	109·47°	1·0887	1·3680	1·2903	
	180°	—	1·8292	—	2·0040
	180° (with m.-s.)	—	2·19	—	2·27
Fe	90°	1·0000	1·7274	1·1852	
	180°	2·0000	2·7603	—	
	180° (with m.-s.)	2·02	2·77	—	
Co	180°		4·0000		

For 180° boundaries in both nickel and iron the angle λ is that between the projections of $\mathbf{I}_0$ and an easy direction on a plane whose normal is the initial direction of $\mathbf{I}_0$ (see § 3). In Table 6.2 the values of b/b_0 for $\lambda=1°$ are included in brackets below the infinite values for $\lambda=0°$.

Table 6.2.

Reduced widths, b/b_0, of domain boundaries in ferromagnetics. b, effective boundary width.
For estimated values of b_0 see text.

Type of boundary		Direction of normal to boundary			
		[001]	[110]	[111]	$[\bar{1}\bar{1}2]$
Ni	70·53°	3·8476	4·2642	—	
	109·47°	∞	3·3093	3·8476	
	180°	—	∞ (15·4029)	—	7·9161
	180° (with m.-s.)	—	7·91	—	4·45
Fe	90°	3·1416	3·9738	3·1416	
	180°	∞ (12·6258)	5·6018		
	180° (with m.-s.)	10·87	5·59		
Co	180°		3·1416		

The effect of introducing a magneto-elastic energy term is most important in nickel, less important in iron, and of least importance in cobalt. Values of γ_0 and b_0 are given by Stoner (1950, table 3.2), who also discusses the uncertainty of these estimates due to uncertainties in the detailed treatment of exchange interaction effects. For nickel, iron and cobalt γ_0, measured in erg cm^{-2}, is 0·14, 0·62 and 2·05; b_0, measured in 10^{-8} cm, is 260, 130 and 50. (Néel (1944) obtains for iron values of 0·7 and 165 for γ_0 and b_0 respectively; Kittel (1949), using the method employed by Lifshitz (1944), in which the exchange energy is calculated from the coefficient of $T^{3/2}$ in Bloch's law of approach to saturation at low temperatures, obtains corresponding values of 0·9 and 230.) Thus the energies, γ erg cm^{-2}, of 180° boundaries in nickel, iron and cobalt (taking $\lambda=0°$ for nickel and iron, and without neglecting magnetostriction) are 0·31, 1·25 and 8·20; the widths, b, in 10^{-8} cm are 2060, 1410 and 160 respectively. These comparisons have a bearing on the character and relative ease of production of powder patterns and on the sizes of domains in the three ferromagnetics (see Stoner 1950). For nickel, iron and cobalt, the boundary widths, at room temperature, are, in general, small compared with the linear dimensions of the domains. If the boundary widths are of the same order of magnitude as the domain size the ferromagnetic can no longer be regarded as being magnetized along some easy direction over most of its volume. This effect may be important in many alloys, and also in the ferromagnetic elements themselves at higher temperatures.

I am indebted to Professor E. C. Stoner, F.R.S., for suggesting the problem and for his guidance and constant encouragement; to Dr. P. Rhodes for helpful discussions and constructive criticism; and to the Department of Scientific and Industrial Research for a grant.

REFERENCES.

BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus*. (Berlin : Springer). (Photo-lithoprint Reproduction, 1943, Ann Arbor, Michigan : Edwards Brothers).

BITTER, F., 1937, *Introduction to Ferromagnetism*. (New York : McGraw-Hill).

BLOCH, F., 1932, *Z. Phys.*, **74**, 295.

BOZORTH, R. M., MASON, W. P., MCSKIMIN, H. J., and WALKER, J. G., 1949, *Phys. Rev.*, **75**, 1954.

COMRIE, L. J., 1949, *Chambers's Six-Figure Mathematical Tables, Vol. 2, Natural Values*. (London and Edinburgh : Chambers).

KITTEL, C., 1949, *Rev. Mod. Phys.*, **21**, 541.

LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet.*, **8**, 153.

LIFSHITZ, E., 1944, *J. Phys. U.S.S.R.*, **8**, 337.

NÉEL, L., 1944, *Cahiers de Phys.*, **25**, 1.

STONER, E. C., 1948, *Phys. Soc. Rep. Progr. Phys.*, **11**, 43 ; 1950, *Ibid.*, **13**, 83.

LXXII *The Statistical Analysis of Electrical Noise.*

By D. K. C. MacDonald,
Clarendon Laboratory, Oxford *.

[Received May 19, 1950.]

Abstract.

Fundamental criticism is made of two recent papers by Meltzer on electrical noise on general and particular grounds. Correction of the errors in the analysis also provides an opportunity for a discussion of the fundamental relation between a stochastic equation and the corresponding frequency spectrum of a fluctuating variable.. An example taken from the field of classical Brownian movement clarifies the concept of the observable velocity of a particle.

Two recent papers by Meltzer (1949, 1950) in this Journal on electrical current fluctuations (noise) present much cause for criticism. The first, (A), implies that a proof has been given of a fundamental physical identity between thermal and shot noise. But in brief the paper simply reiterates the well-known fact that any purely random phenomenon, which may be idealized as a Markoff process, obeys a certain rather simple statistical law. Membership of the same statistical class, however, is remote from an essential physical identity of two phenomena, and the suggestion that a serious proof of this has been given is at least rather misleading. The second paper, (B), however, takes up the problem of current noise in metallic resistors and uses analysis which is invalid in fundamental, but elementary, respects; the ensuing results and comparison with the experimental work of Bernamont are consequently fallacious.

We first consider the specific errors of the second paper and then take the opportunity of a more general discussion (relevant to both papers) of the statistics and frequency spectrum of fluctuations. An example taken from the field of classical Brownian movement clarifies the concept of the observable velocity of a particle.

It is simplest in considering (B) to present afresh in correct form Meltzer's analysis, adhering as far as possible to his notation. Consider some cross-section of a conductor and let us observe the charge crossing this during a set of time intervals of (arbitrary) length, τ †. Let n be the total number of charges e_i, of magnitude e, crossing in any particular member of the set. This number will of course not be constant, because

---

* Communicated by the Author.

† We prefer to use τ in place of Meltzer's t.

of the very presence of fluctuations, but will vary from interval. Let n_1 be the number crossing in the "positive" direction and n_2 in the "negative" direction, and let the respective probabilities be p and q. Thus denoting the *mean* values of the three statistical variables n, n_1 and n_2 by N, N_1 and N_2 we have:

$$N_1=pN\,;\ N_2=qN$$

and the mean current, I, is given by:

$$I=(p-q)\frac{eN}{\tau} \qquad \text{(1)}$$

Then, following Meltzer, we set:

$$Q=\sum_{i=1}^{n} e_i=(n_1-n_2)e \qquad \text{(2)}$$

$$\therefore\ \overline{Q}=(\bar{n}_1-\bar{n}_2)e=Ne(p-q)=I\tau$$

and
$$Q^2=\sum_i e_i^2+\sum_{i\neq j} e_ie_j \qquad \text{(3)}$$

$$\therefore\ \overline{Q^2}=\bar{n}_1e^2+\bar{n}_2e^2+\overline{n_1(n_1-1)}e^2+\overline{n_2(n_2-1)}e^2-2\overline{n_1n_2}e^2 \qquad \text{(4)}$$

$$=(N_1+N_2)e^2+(\overline{n_1^2}-N_1)e^2+(\overline{n_2^2}-N_2)e^2-2\overline{n_1n_2}e^2.$$

Now let

$$n_1=N_1+\delta_1\,;\ n_2=N_2+\delta_2$$

$$\therefore\ \overline{n_1^2}=N_1^2+\overline{\delta_1^2}\,;\ \overline{n_2^2}=N_2^2+\overline{\delta_2^2}\,;\ \overline{n_1n_2}=N_1N_2+\overline{\delta_1\delta_2}$$

and
$$\overline{Q^2}=Ne^2+(N_1-N_2)^2e^2+(\overline{\delta_1^2}+\overline{\delta_2^2}-2\overline{\delta_1\delta_2}-(N_1+N_2))e^2 \qquad \text{(5 a)}$$

$$=Ne^2+N^2(p-q)^2e^2+(\overline{\Delta^2}-N)e^2, \qquad \text{(5 b)}$$

if we set $\Delta\equiv\delta_1-\delta_2$.

This should be compared with Meltzer (B), equation (5). Meltzer has throughout made no distinction between average and mean square values, etc., of the relevant statistical variables, which appears a rather serious error in a paper intended to deal with fluctuations.

The achievement of further real progress depends now very naturally on the assumption of some *physical* model for the system considered. The only assumption of a general nature that we can essay *here* is that of an overall equilibrium distribution:

$$\overline{\Delta^2}=N \qquad \text{(6 a)}$$

or more specifically to assume:

$$\overline{\delta_1^2}=N_1\,;\ \overline{\delta_2^2}=N_2\,;\ \overline{\delta_1\delta_2}=0 \qquad \text{(6 b)}$$

both yielding:

$$\overline{Q^2}=Ne^2+N^2(p-q)^2e^2 \qquad \text{(7)}$$

and using (1):

$$\overline{Q^2}=Ne^2+I^2\tau^2. \qquad (8)$$

Using finally Einstein's classic result (Einstein (1906)) that for a resistor, R, in thermal equilibrium at temperature T, the effective fluctuation in charge ($\overline{\Sigma e_i^2}$ after equation (3)) is given by $(2kT/R)\tau$, we may write:

$$\overline{Q^2}=\left(\frac{2kT}{R}\right)\tau+I^2\tau^2. \qquad (9)$$

This is of course just Meltzer's starting point, (B) equation (1), from which no amount of algebra by itself can possibly derive any essentially new expression for the fluctuation involved. By the very fact of writing down this equation, Meltzer has implicitly assumed a complete knowledge of the statistics of the fluctuation in the presence of the term in τ, representative of a purely random component, and the term in τ^2, representative of a pure "d.c." contribution. The simple addition of the two terms embodies immediately the assumption that these are entirely independent—or physically, to sum up, that the passage of a direct current I *has had no effect whatsoever on the fluctuations to be observed.*

In particular, then, all equations subsequent to (5) in Meltzer (B) are spurious and consequently the suggested comparison with the experiments of Bernamont (1937) is invalid.

Let us remark that it may prove possible to set up a phenomenological theory of current noise in thin metallic films and semiconductors proceeding from (our) equations (5) by making appropriate assumptions about the last term (in either form) in any particular physical model. In any case, however, great care will be necessary in formulating and interpreting an analysis of this type since the current observed at the terminals of a resistance, say, is composed of the contributions of *all* electrons in the body and a fluctuating current may well be induced by motion of remote charge without electrons actually crossing locally a specific cross-section of the body. It is in this sense that a rigid interpretation of N may be entirely misleading and, in particular, Meltzer's "equivalent shot-effect current" defined by $I_0=kT/Re$ in (A) has no very direct physical significance.

A general feature which perhaps is not sufficiently widely appreciated is the complete, immediate equivalence of a statistical equation such as (9) above to the statement of the current fluctuations in terms of a frequency spectrum. If we are provided with the auto-correlation function $\overline{y(t).y(t+\tau)}^{t}$ of a stationary statistical variable, $y(t)$, then it should now be well known that the theorem of Wiener and Khintchine provides an immediate transformation to the corresponding (power) frequency spectrum. In addition, however, as we have shown (MacDonald (1949)) a knowledge of the average mean-square deviation of a variable over any arbitrary time-interval τ also leads directly to the frequency spectrum:

this concept is sometimes more natural than the auto-correlation function which deals with the *instantaneous* values of the variable concerned. Thus equation (9) is in fact precisely equivalent to writing :

$$\overline{I_f^2}=\frac{4}{R}kT\Delta f+I^2\,.\,\delta(f)\quad.\;.\;.\;.\;.\;.\;.\quad(10)$$

(where $\delta(f)$ is a Dirac function " at " zero frequency, representative of the *direct current*, I, such that $\int_0^\infty\delta(f)df=1$).

An interesting application of these ideas arises in the original extended equations of Brownian movement due to Langevin (1908) and others (*e.g.* Fürth (1920)).

If x be the displacement of a free Brownian particle, of mass m, in a medium of viscous resistance λ at temperature θ, then in time τ :

$$\overline{\delta x^2}=\frac{2k\theta}{\lambda}\{\tau-T(1-\epsilon^{-\tau/T})\},\quad\text{where}\quad T=m/\lambda,\quad.\;.\;.\;.\quad(11)$$

thus the mean square deviation of the velocity averaged over the interval τ is

$$\overline{\delta v_\tau^2}\equiv\frac{\overline{\delta x^2}}{\tau^2}=\frac{2k\theta}{\lambda}\left\{\frac{1}{\tau}-\frac{T}{\tau^2}(1-\epsilon^{-\tau/T})\right\}.\quad.\;.\;.\;.\;.\quad(12)$$

We can then immediately, as pointed out above, derive the frequency spectrum of the true " instantaneous " velocity :

$$\overline{\delta v_f^2}=\frac{4k\theta}{\lambda}\left\{\frac{df}{1+(2\pi fT)^2}\right\}.\quad.\;.\;.\;.\;.\;.\;.\;.\quad(13)$$

By integration over all f from 0 to ∞, we have

$$\overline{\delta v^2}=\frac{4k\theta}{2\pi\lambda T}\cdot\frac{\pi}{2}=\frac{k\theta}{m}\quad.\;.\;.\;.\;.\;.\;.\;.\quad(14)$$

in agreement with the classical equipartition formula of Einstein, while (13), on the other hand, shows immediately that, unless our measuring instruments respond to, or resolve, events of natural frequency considerably greater than 1/T, we shall not in fact observe the " true " Brownian movement velocity as given by (14) (*cf.* Ramsay (1892), Exner (1900), Svedberg (1906)). If, however, the limiting resolution time of our instrument is, say, T_1 then (13) enables us to *determine* the value of $\overline{\delta v^2_{\text{obs}}}$ which should in fact be observed. Thus :

$$\begin{aligned}\overline{\delta v^2_{\text{obs}}}&=\frac{4k\theta}{\lambda}\int_0^{1/T_1}\frac{df}{(1+(2\pi fT)^2)}\\&=\frac{2k\theta}{\pi m}\left(\arctan\frac{2\pi T}{T_1}\right),\quad.\;.\;.\;.\;.\;.\;.\quad(15\,a)\end{aligned}$$

or, if $T\ll T_1$,

$$\overline{\delta v^2_{\text{obs}}}\approx\frac{4k\theta}{m}\left(\frac{T}{T_1}\right).\quad.\;.\;.\;.\;.\;.\;.\;.\;.\;.\quad(15\,b)$$

In examining the early literature to find measurements for comparison with equations (15), I was unable to find entirely satisfactory data, generally because no very precise record of T_1 was kept. However, those of F. M. Exner (*loc. cit.*) enable a good comparison to be made. Since $T=m/\lambda$ and $\lambda=6\pi\eta a$, for a small spherical particle, of radius a, observed in a medium of viscosity η, we have:

$$\overline{\delta v^2_{\text{obs}}}\approx\frac{4k\theta}{6\pi\eta a T_1} \qquad \text{(16)}$$

It appears that Exner followed visually the track of the particle continuously, in so far as he was able. It therefore appears reasonable to take $T_1\sim 0{\cdot}1$ second, with an upper limit certainly of 1 second.

Now,

$$a=3{\cdot}5\times 10^{-5}\text{ cm.}$$
$$\eta\approx 10^{-2}\text{ poises.}$$
$$\theta=293^\circ\text{ K. }(20^\circ\text{ C.}).$$

Whence $\sqrt{(\overline{\delta v^2_{\text{obs}}})}\approx 5\times 10^{-4}$ cm./sec. (with a lower limit of $1{\cdot}6\times 10^{-4}$ cm./sec.). The velocity observed by Exner at this temperature was in fact $3{\cdot}2\times 10^{-4}$ cm./sec., and the agreement is thus very satisfactory.

This analysis appears also of interest because it has been stated in the past that the velocity of a Brownian particle as observed visually has only a subjective significance.

I should like to thank Professor M. H. L. Pryce and Mr. D. F. Gibbs for helpful discussion.

References.

BERNAMONT, J., 1937, *Ann. de. Phys.*, **7**, 71.
EINSTEIN, A., 1906, *Ann. Phys., Lpz.*, **19**, 371.
EXNER, F. M., 1900, *Ann. Phys., Lpz.*, **2**, 843.
FÜRTH, R., 1920, *Zs. f. Phys.*, **2**, 244.
LANGEVIN, P., 1908, *Comptes Rendues*, **146**, 530.
MACDONALD, D. K. C., 1949, *Phil. Mag.*, **40**, 561.
MELTZER, B., 1949, *Phil. Mag.*, **40**, 1224; 1950, *Ibid.*, **41**, 393.
RAMSAY, W., 1892, *Chem. News*, **65**, 90.
SVEDBERG, TH., 1906, *Zs. f. Elektrochem.*, **12**, 853, 909.

LXXIII. *The Origin of Single Cosmic Ray Protons.*

By S. Lattimore,
Imperial College of Science and Technology, London *.

[Received June 2, 1950.]

Abstract.

The frequency of production, and the angular and energy distributions of single cosmic ray protons produced in photographic emulsions exposed at 3650 m. are given. It is shown that, after corrections for protons emitted from stars have been made, the results agree with those given in a previous paper (Lattimore 1949).

The rates of production of fast protons in paraffin wax, carbon, aluminium, and lead are deduced from measurements of the number of protons observed to cross 100μ C.2 emulsions surrounded completely by these materials. From the results for paraffin wax and carbon, the rate of production in hydrogen is also deduced.

Introduction.

In a previous paper (Lattimore 1949) it has been shown that the number of proton tracks observed crossing 100μ C.2 emulsion is several times greater than the number to be expected if these tracks were due solely to protons from cosmic ray stars in the surrounding material. From the number of such tracks, and the number of protons observed to end in the emulsion, it was deduced that 62 protons were produced/c.c/day in emulsion at 3650 m., and that their energy of production was 30 MeV. or more. In the calculation of the number of protons crossing the emulsion, star protons of energies greater than 30 MeV. were ignored. The number of such star protons was therefore included in the value of 63/c./c./day.

It must be emphasized that at the time of these previous experiments, emulsions sensitive to tracks of minimum or near minimum ionization were not available. The C.2 emulsion was not of sufficient sensitivity for protons with energies of 30 MeV. or greater to give recognizable tracks. Now, however, emulsions sensitive to minimum ionization tracks are available, and it is possible to study the actual beginnings of the single proton tracks. The present paper is concerned with the origin of these single protons and their angular and energy distributions. The relative numbers of protons produced in various materials is also given.

---

* Communicated by Sir George Thomson, F.R.S.

PART A.

1. *Frequency of Production.*

Ilford G.5. 200μ emulsion on glass were used. The emulsion was searched twice, using a ×95 oil immersion objective. A total of 89 fast protons, other than those produced by stars, were observed to start within the emulsion. Of these, 74 occurred singly. Five cases were observed where two fast protons were produced together. Of these, two occur near the surface, and may be coincidences. One case was observed of what may be a proton of energy greater than 100 MeV. either producing a proton of 60 MeV. and vanishing itself, or losing considerable energy in passing through a nucleus, without producing visible tracks. Two cases were found where a fast proton apparently produces a second proton, and escapes itself. The value of 25 ± 2/c.c./day is obtained for the rate of production in emulsion of these fast protons at 3650 m. In addition to these protons, we now observe many fast protons emitted from stars. These were not observed in the C.2 emulsions, in general, because of the low sensitivity. In Table I. is shown the number of protons of energy greater than 30 MeV. associated with stars.

TABLE I.

Star Size	Mean No. of Protons of Energy greater than 30 MeV. per star	No. of Protons of Energy greater than 30 MeV. produced/c.c./day
2, 3, or 4 prongs	1·1	20·5
4 prongs	1·26	7·5

This gives a total of 28 protons of energy greater than 30 MeV. produced/c.c./day in emulsion at 3650 m. from stars.

Considering, therefore, both the protons produced singly, or in processes other than normal star production, and those protons of energy greater than 30 MeV. produced in stars, we find a total rate of production of 53/c.c./day in emulsion at 3650 m. These are protons actually observed to start in the emulsion.

2. *Angular and Energy Distribution.*

The angles between the projection of the proton tracks on the surface of the emulsion and the vertical were measured by means of a protractor attached to one of the eyepieces of the microscope. The angles of dip were also measured, by means of a graduated fine focusing adjustment. This was calibrated by measuring the thickness of the emulsion, which was assumed to be 200 microns. The lengths of the projections of the tracks on the surface of the emulsion were measured, and the number of grains along the tracks. From these four measurements we could calculate the true angle between the proton and the vertical, and also true grain density. Not all of the tracks found were suitable for measurement, either because of shear in the emulsion, or because they dipped too steeply for us to perform a reliable grain count.

A grain density-energy curve for protons was constructed by measuring grain densities along ρ mesons which stopped in the emulsion, and using

Energy in MeV.

the relation $dN/dR=e^2f(\beta)$, where dN/dR is the grain density, "e" the charge of the particle, and $f(\beta)$ some function of the velocity of the particle.

Fig. 1 *a*.

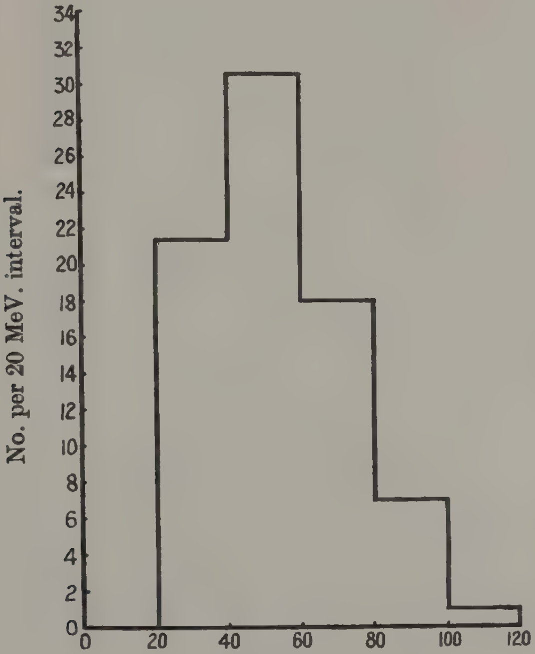
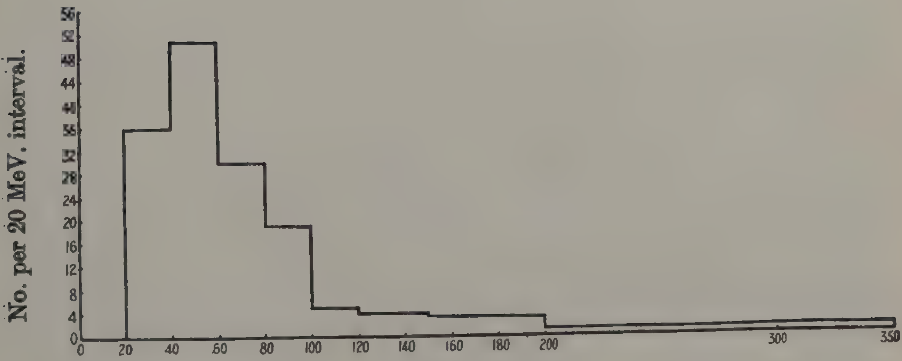


Fig 1 *b*.

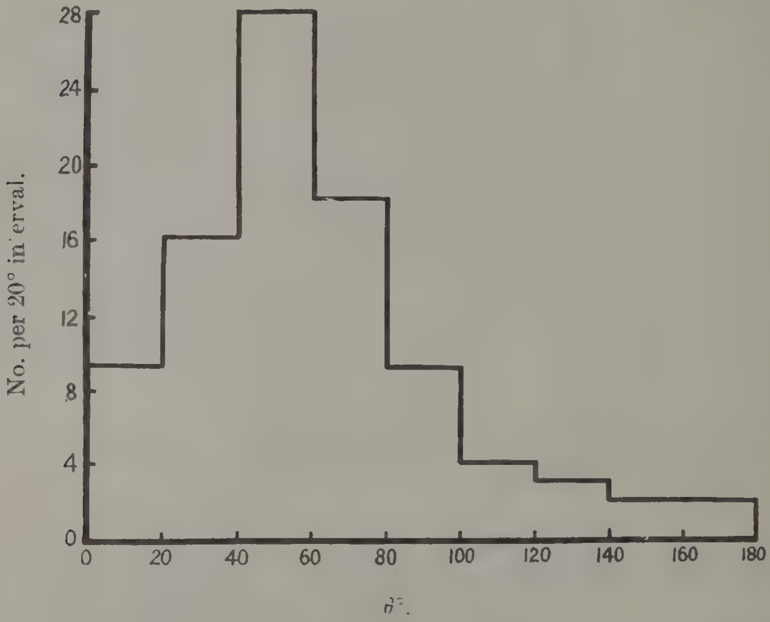


Energy in MeV.

The energy distribution of those protons produced other than in stars is shown in fig. 1*a*. It has a maximum between 40 and 60 MeV., and falls to zero at 120 MeV. The energy distribution for both protons from stars and protons produced singly is shown in fig. 1*b*.

The distribution shown in fig. 1 *a* falls to zero at an energy of 20 MeV. Single protons are produced with energies below 20 MeV. (see previous paper), but most of these have very low energies (5 MeV.) and it is probable that these are true single prong stars. The cut-off at 20 MeV. is therefore somewhat arbitrary.

Fig 2 *a*.



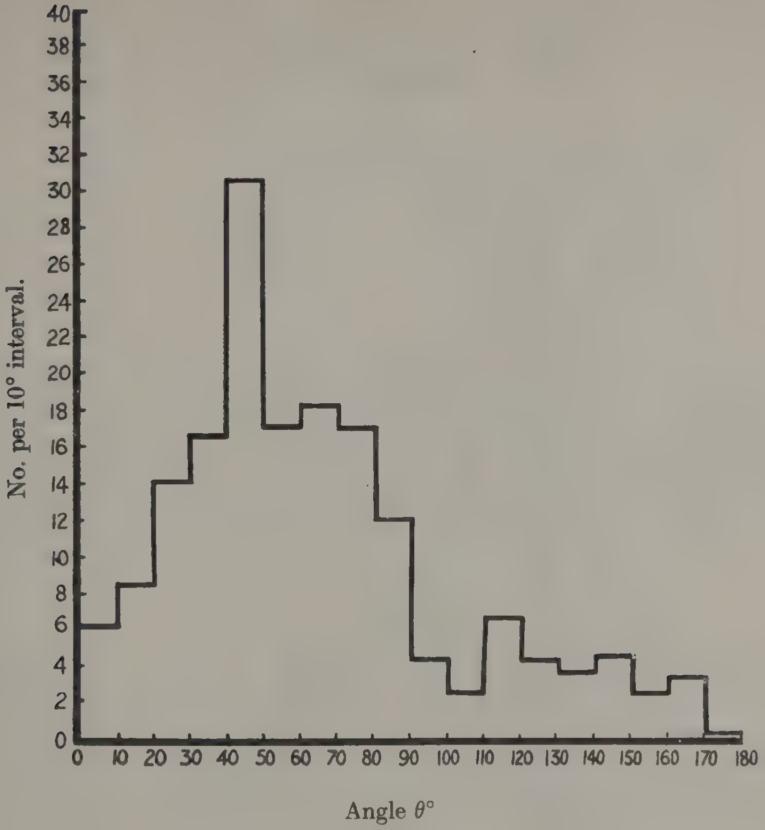
In fig. 2*a* is shown the angular distribution of the single protons, of energy greater than 20 MeV., at production, and in fig. 2*b*, the angular distribution of all fast protons produced, *i. e.* both those produced in stars and those in other processes. In these figures θ is the angle between the track and the vertical.

3. *Comparison with Previous Results.*

In a previous paper (Lattimore 1949) it was deduced that 62 fast protons were produced c.c./day in emulsion at 3650 m. This value was deduced from the number of proton tracks observed to cross the emulsion, and from an estimate of the sensitivity of the emulsion used. It is clearly difficult to assess this latter quantity with any accuracy, and it is probable that the sensitivity was underestimated in these earlier experiments.

In fig. 3 is shown the angular distribution of all fast protons produced. $N(\theta)$ represents the number of protons emitted/unit solid angle at an angle of $\theta°$ or $(180-\theta)°$ to the vertical. The full line is the distribution obtained in the previous paper (fig 4*b*). The agreement is reasonably good, although the statistics are poor.

Fig. 2*b*.



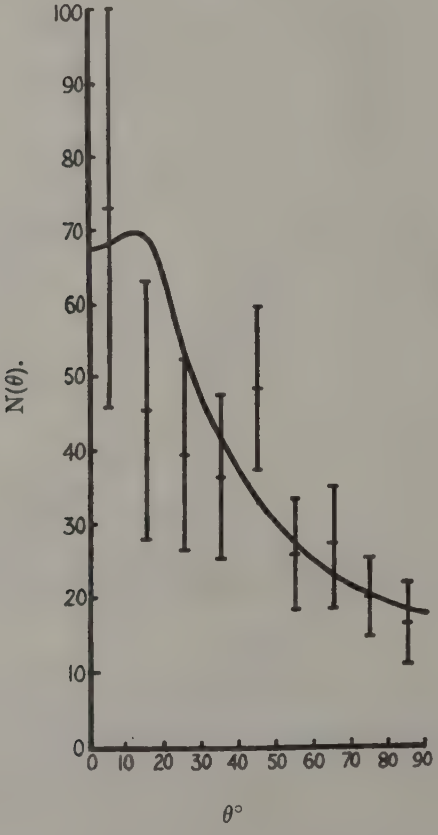
Part B.

Variation of Frequency of Production with Material.

For this part of the experiment, it was necessary to revert to the original method of counting, single protons which crossed the emulsion, and which were therefore produced in the surrounding material. 100μ C.2 emulsions on film were used and these were surrounded on all sides by the various materials and exposed horizontally, emulsion uppermost, at 3650 m. C.2 emulsions were used in preference to G.5 because the latter records all ionizing particles, including fast mesons and electrons. It was found that the number of protons crossing the emulsion was independent of the surrounding material, in agreement with results of Perkins (1947) and Morand and Cuër (1949).

In Table II. are shown the number of protons observed to cross the emulsion with the different surrounding materials. Also shown are the calculated rate of production in the various materials. For the calculation of the rate of production, it is necessary to know the sensitivity of the emulsion, by which we mean the energy of the fastest proton that can be observed to cross the emulsion (see previous paper). This was taken as 35 MeV., a value which brings the rate of production given in our previous paper (Lattimore 1949) into agreement with the rates found in emulsion in the present experiment. We can then find the rates of production in the various materials by using the formula given in the previous paper. The ranges were calculated from the equation

Fig. 3.



$$dE/dx = 4{\cdot}15\rho(Z/A)\mu_e/\beta^2\left[\log\frac{\beta^2}{(1-\beta^2)I}-\beta^2\right]$$

(Bethe 1937), where ρ, Z, and A are respectively the density, atomic number and atomic weight of the stopping material, μ_e the mass of the electron, β the velocity of a proton of energy E, and x the range of the proton in gms./cm.^{2}. I is the ionization potential of the stopping material (see Table II.). This equation was integrated graphically. Since the calculated absolute values for the rate of production depend upon the value taken for the sensitivity of the emulsion used, they are perhaps open to doubt, but the relative values should be approximately correct.

TABLE II.

Material	No. crossing emulsion Arbitrary units	I(eV)	No. produced/gm/day
Pb.	38·7 ± 2	1150	6·5 ± 0·5
Al.	37·4 ± 2	175	9·65 ± 0·7
C.	40·4 ± 2	81	15·1 ± 1·1
CH.	42·7 ± 2		18·6 ± 1·4
H.		13·5	39·6 ± 12

Discussion of Errors. The principal error involved, other than purely statistical errors, is introduced by taking the emulsion thickness as 200 μ. The thickness was unfortunately not measured before development. Measurements on two further batches give errors of 6 per cent and 8 per cent in the thickness. Any error will affect the angular distributions only slightly by altering the angle of dip, but affects the number of single tracks produced/c.c./day directly.

CONCLUSIONS.

This work has confirmed our earlier conclusions given in the previous paper, that protons are produced singly, and that they are mainly directed downwards.

It would seem reasonable to suppose that these protons are produced by the same primaries as produce stars, since the number both of stars and single tracks vary in the same way under ice (Harding *et al.* 1949) and under air (Lattimore unpublished). It follows that any theory of single proton production must take account of star production also. Attempts are being made to produce such a theory, and any results will be published later.

REFERENCES.

HARDING, 1949, *Phil. Mag.*, **40**, 530.
LATTIMORE, 1949, *Phil. Mag.*, **40**, 394.
LIVINGSTONE and BETHE, 1937, *Rev. Mod. Phys.*, **9**, 245.
MORAND and CÜER, 1949, *Cosmic Radiation* (London : Butterworths).
PERKINS, 1947, *Nature, Lond.*, **160**, 707.

LXXIV. *The Control of a Wilson Cloud Chamber by means of an Internal Counter.*

By A. L. Hodson, A. Loria[*] and N. V. Ryder[†].
The Physical Laboratories, The University, Manchester[‡].

[Received June 8, 1950.]

[Plates XXV.–XXVIII.]

Abstract.

A technique is described for controlling a Wilson cloud chamber by an internal counter whose cathode consists of a hexagonal array of rods and whose sensitive volume is not separated from the main cloud chamber volume by any wall and in fact forms part of the usable chamber volume. The counter has no stopping power beyond that of the gas unless the incident particle chances to hit one of the rods. Events in which heavily ionizing particles are produced in the cloud chamber and events in which low energy particles traverse the chamber can thus be photographed. Discrimination between single fast particles or events of high ionization can be made at will. Altogether 26 slow particles (excluding electrons) have been photographed—corresponding to a vertical intensity of $9{\cdot}3\times10^{-5}$ slow particles cm.$^{-2}$ min.$^{-1}$ sterad.$^{-1}$ at sea level. Details of a proposed arrangement for photographing nuclear stars in the gas of the chamber are given. The technique is applicable to the study of the discharge phenomena of counters and some preliminary photographs of these phenomena are presented. The counter is quite stable and the technique of operation is not unduly difficult.

§1. Introduction.

In many cloud chamber experiments it would be an advantage if the chamber could be triggered on the occurrence of a heavily ionizing event. A proportional counter operated within the chamber offers a means of detecting such events, but for the observations of processes involving or caused by very low energy particles the absorption in the wall of the counter is an important factor. The possibility of employing a counter in which this absorption was reduced to a minimum by means of a cathode assembly consisting of rods was investigated by Hodson and Page in 1948 with a view to detecting and photographing nuclear stars occurring in the gas of a cloud chamber[||]. The idea of using a cathode of this form was obtained from the work of McCusker (1942) who had used such an arrangement for Geiger-Müller counters in glass envelopes. After experimenting with various rod assemblies McCusker showed that a hexagonal arrangement of rods was the most satisfactory. Hodson and Loria (1949) succeeded in controlling a Wilson cloud chamber by an internal counter having a cathode of this type but whose sensitive volume was not separated from the main chamber volume by any envelope and in fact formed part of the usable chamber volume.

---

[*] Now at the Università di Padova, Italy.

[†] On leave from Victoria University College, New Zealand.

[‡] Communicated by Professor P. Ms. Blackett, F.R.S.

[||] After this initial work, attention was drawn to a proposal of Cohen to use a cloud chamber as an ionization chamber for this purpose (Cohen 1948). The success of this method has since been reported (Cohen 1949a, b).

This paper gives details and progress of this technique, together with some preliminary results obtained.

§2. Experimental Details.

(a) *Constructional details of the counter* :

The counter at present employed is 25 cm. in length and 4 cm. in diameter and is illustrated in Pl. XXV. .The cathode rods are of steel, 2 mm. in diameter, supported at each end by a brass framework, the whole cathode assembly being chromium plated. It is mounted across the middle of a cloud chamber of 28 cm. diameter and 9 cm. depth. The anode wire is attached to two metal supports (for details see fig. 1) passing through holes in the glass cylinder of the chamber, the holes then being sealed with small rubber gaskets. The compressibility of the gaskets also provides a useful means for the final tensioning of the wire. Both tungsten and platinum wires (diameter 0·1 mm. and 0·17 mm.) have been used successfully, but, for reasons to be given later, platinum wires are preferable. In future applications it is proposed to use cathode rods of smaller diameter with surfaces blackened to reduce the amount of light scattered.

Fig. 1.

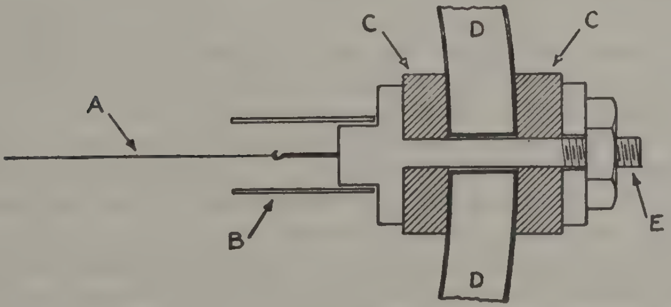


Diagram of wire mounting.

A. Anode wire.
B. Shield·
C. Rubber gasket.
D. Glass cylinder.
E. Threaded support.

(b) *Chamber filling* :

In normal cloud chamber operation advantage is taken of the readiness with which electrons produced by ionizing particles attach themselves to oxygen molecules in the gas to form heavy negative ions. In counters, however, the production of gas multiplication depends on the electrons remaining freely mobile, and electron attachment must be avoided. For this reason precautions must be taken to exclude oxygen and water vapour. The cloud chamber is therefore filled with commercial argon (stated to be of 99·8 per cent purity and oxygen-free) and saturated with absolute ethyl alcohol vapour—a mixture which gives good cloud chamber tracks and at the same time is a satisfactory filling for the counter. As it is not possible to evacuate the chamber, the air initially present is removed by partial extraction with argon : argon is admitted to a pressure of 1·5 atmospheres and the pressure is then released to one atmosphere. This procedure is repeated at least ten times and the chamber then finally filled with argon to the required pressure of about 118 cm.Hg in the compressed position.

If there is no appreciable negative ion formation the electrons produced along the path of an ionizing particle are rapidly swept away by the general clearing field applied across the chamber. The tracks observed when the cloud chamber is expanded thus arise from positive-ion condensation only.

It is found that if no further precautions are taken the pulse size of the counter decreases steadily over a period of several days after the cloud chamber has been filled—presumably due to the slow release of oxygen and water vapour from the walls and velvet background of the chamber and to the slow diffusion of air into the chamber through the rubber diaphragm. The increasing amount of electron attachment which thus occurs results in an undesirable doubling of the tracks, showing that condensation is now occurring on both positive and negative ions. (A similar difficulty has been experienced by Cohen (1949 b) with his combination cloud- and ion-chamber). The *status quo* can be restored temporarily by reflushing the chamber with argon and for the present purposes it has proved sufficient to do this once a week. However, a method of continuously removing oxygen and water vapour is essential for a permanent and reproducible pulse size in the counter, and we are now investigating various possible methods.

(c) *Operation of the counter* :

The counter will operate satisfactorily either in the proportional region or in the Geiger region. With the pressures used (118 cm. of mercury) and a wire diameter of 0·17 mm., suitable proportional operation is achieved over the range 1950–2300 volts, while the threshold potential for the Geiger region is about 2500 volts.

The cathode rods of the counter are earthed and the anode wire is connected to a positive high voltage supply through a 2 megohm resistance.

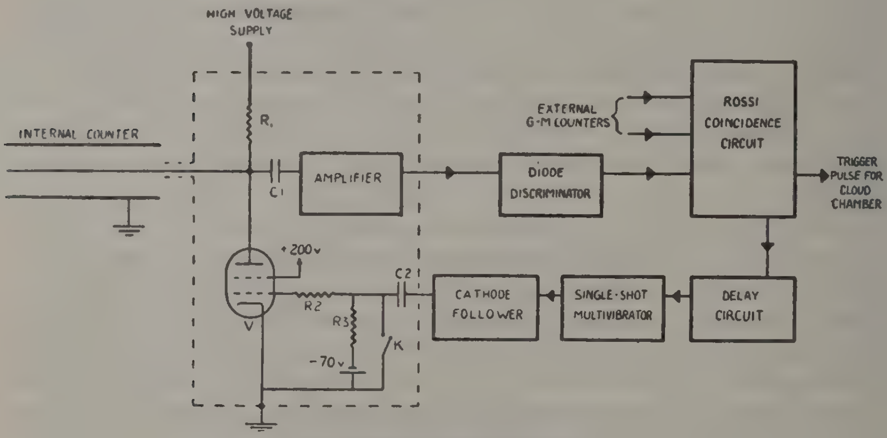
It will be seen later that the most suitable method for detecting particles inside a cloud chamber is to operate in the proportional region, using a fairly low gas multiplication and high external (valve) amplification. In this mode of operation the negative pulses obtained from the wire are amplified by a high gain linear amplifier (gain variable between 500 and $3 \cdot 10^5$) incorporating a biased diode discriminator at the output stage, which, in conjunction with the gain setting, determines the minimum size of pulse that will be accepted.

Particular care is necessary to avoid spurious pulses at the input stage of the amplifier. Any appreciable leakage from high voltage points associated with the anode wire will give rise to spurious pulses, but by mounting the anode wire directly on the glass cylinder of the chamber so that the leakage path is made as long as possible and by using polystyrene supports for all high voltage components, leakage is readily reduced to a negligible level. A silicone varnish (trade name " Aquatect ") has proved very effective in reducing residual leakage across the surface of the glass cylinder. These initial precautions being taken, the counter has been operated continuously for many weeks without any trouble due to spurious pulses.

For these initial investigations the counter has been operated in coincidence with two Geiger counters placed above the chamber in a vertical telescope arrangement. The output pulse from a threefold Rossi coincidence circuit is used to expand the cloud chamber, and within a few microseconds (*q. v.*) of a coincidence being detected the high voltage on the wire is removed electronically. The voltage is thus removed before the positive ions along the path of the incident particle have had time to move any appreciable distance and the track is therefore visible and undistorted even within the sensitive volume of the counter (Photograph (*a*), Pl. XXVI.). The small " bead " of ionization observed on the wire near to the track of the incident particle consists of the positive ions formed by the gas multiplication which occurs near the wire during the discharge of the counter. By using high values of the amplifier gain and quenching rapidly it is, in fact, possible to remove the voltage on the wire before the process of gas multiplication has reached completion—the " bead " observed is then correspondingly smaller and may even be too small to be visible on the cloud chamber photograph. However, it is often desirable to know which particle discharged the counter and this is possible if a track can be associated with a " bead " on the wire. For this reason the removal of the anode voltage is usually delayed by approximately 20 microseconds ; this allows the discharge to develop until it is quenched by a space-charge reduction of the field near the wire (Stever 1942), and a " bead " of reasonable size is observed.

The columns of ionization marked W in Photographs (*a*) and (*b*), Pl. XXVI., and also visible in the subsequent photographs arise as a result of surface irregularities of the wire. If the wire is not smooth or if particles of dust adhere to it an " electric wind " or corona phenomenon produces continuous streams of positive ions moving away from the wire and occurring at the same points in successive photographs. This phenomenon does not interfere directly with the normal operation of the counter but the streamers are rather undesirable from the point of view of clear cloud chamber pictures. Fortunately it is possible to avoid most of them by cleaning the wire carefully. The initial cleaning consists merely of washing the wire with alcohol with a fine brush. Platinum wires are preferable to tungsten as they have a smoother surface and may be safely glowed in air after closing the chamber as the final stage of the cleaning process. The " electric winds " have been investigated in some detail and the exact mechanism of their production will be discussed in a subsequent paper. They are more pronounced at higher voltages on the wire.

Fig. 2.


HIGH VOLTAGE SUPPLY
R1
INTERNAL COUNTER
C1
AMPLIFIER
+200v
C2
R2
R3
V
K
-70v
EXTERNAL G-M COUNTERS
ROSSI COINCIDENCE CIRCUIT
DIODE DISCRIMINATOR
TRIGGER PULSE FOR CLOUD CHAMBER
CATHODE FOLLOWER
SINGLE-SHOT MULTIVIBRATOR
DELAY CIRCUIT

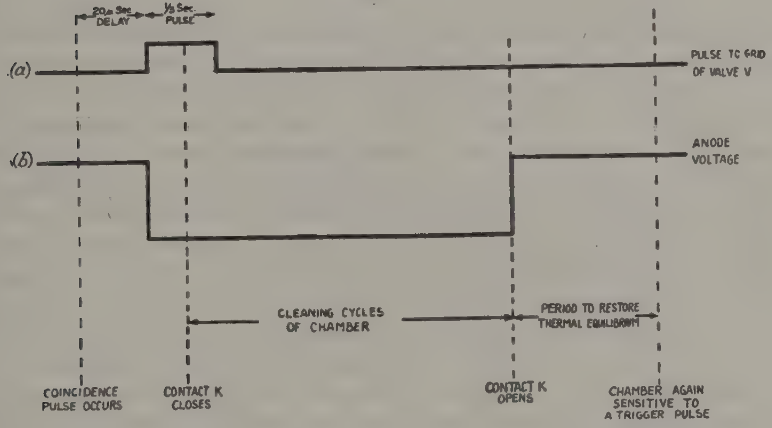
Schematic circuit diagram.

R_1	2 m.Ω.	2 watt.
R_2	30 k.Ω.	½ watt.
R_3	1 m.Ω.	1 watt.
C_1	100μμF.	High voltage.
C_2	0·5 μF.	
Valve V.	CV73 (M.A.P.).	

The general circuit arrangement used for removing the anode voltage of the counter after a coincidence has been detected is shown diagrammatically in fig. 2. The valve V is normally biased beyond anode current cut-off so that the full positive potential is applied to the wire of the counter. When a coincidence is detected a pulse, delayed by about 20 microseconds, triggers a single-shot multivibrator. This produces a square positive pulse (~1/3 second long) which is applied to the grid of the valve V. Valve V then becomes fully conducting and the anode voltage of the counter is rapidly removed. A few milliseconds after the cloud chamber is triggered the control mechanism of the chamber closes a relay contact K and maintains the grid of valve V at earth potential until the cleaning cycles of the chamber have been completed. The anode voltage of the counter is therefore approximately zero during this period (fig. 3) and the production of fresh ionization by discharges of the counter and by the " electric wind " phenomenon during the cleaning expansions is thereby avoided.

Operation of the counter in the manner described above, *i. e.* in the proportional region with removal of the anode voltage a few microseconds after triggering the cloud chamber, will hereafter be referred to simply as " the normal mode of operation ". Other modes of operation will be described in § 3.

Fig. 3.


20μ Sec.
DELAY
⅓ Sec.
PULSE
(a)
PULSE TO GRID OF VALVE V
(b)
ANODE VOLTAGE
CLEANING CYCLES OF CHAMBER
PERIOD TO RESTORE THERMAL EQUILIBRIUM
COINCIDENCE PULSE OCCURS
CONTACT K CLOSES
CONTACT K OPENS
CHAMBER AGAIN SENSITIVE TO A TRIGGER PULSE

Sequence of operations.

§ 3. THE DISCHARGE PHENOMENA OF PROPORTIONAL AND GEIGER-MULLER COUNTERS.

The technique of operating an open counter in the gas of a cloud chamber is one that is obviously applicable to the study of discharge phenomena in counters. When a counter is operated in the Geiger region the discharge is known to spread along the wire from the point where the particle passed through the counter (Stever 1942). Several determinations of the velocity of this discharge spread have been made and various mechanisms for this process have been suggested. Corson and Wilson (1948) have published a useful review of this work and it has been discussed more recently by Balakrishnan and Craggs (1950). Of special interest also is the transition between the proportional region of a counter and the Geiger region. The present technique is being applied to the study of these problems and an account of this work will be published separately. However, a few photographs will be given here to indicate the possibilities of the method.

If the counter is operated in the proportional region and the cloud chamber is triggered without removing the voltage from the anode wire the positive ions formed near the wire in the discharge of the counter move outwards towards the cathode until condensation begins. Instead of forming a " bead " near the wire as in Photograph A, the positive ions now appear as a "ring" of ions in Photograph (*b*) (Pl. XXVI.). Photographs (*a*) and (*b*) demonstrate directly the extreme localization of the discharge in the proportional region.

Photograph (*c*) (Pl. XXVI.) was taken a ta rather higher anode voltage than Photographs (*a*) and (*b*) and is typical of photographs obtained in the transitional region between the proportional and the Geiger regions. It was obtained by removing the anode voltage about 20 microseconds after triggering the chamber and may be compared with Photograph (*a*). At the higher voltage it is seen that the discharge spreads along the wire in a series of quite characteristic and fairly regularly spaced " jumps ", and the positive ion space charge consists of a number of " beads " along the wire. It has been observed that as the anode voltage is further increased the discharge becomes progressively more continuous along the length of the wire.

From the previous description of the normal mode of operation it is clear that the anode voltage is removed only when expansion of the chamber is about to take place. But bursts of ionization are being produced continually in the counter and although the initial radial movement of the " bead " is very rapid the time for the complete collection of the positive ions on the cathode rods is several milliseconds, owing to the high gas pressure used in the chamber. There is thus a probability that besides the selected particle and its associated " bead " of ionization there will be present tracks and " rings " of ionization due to events which occurred within a certain interval before the principal track. Examples of this may be seen on Photographs (*e*) and (*f*), Pl. XXVII. At low voltages, where the discharge is very localized, the " rings " so formed are narrow and not very intense. At voltages sufficiently high to produce spreading of the discharge along the wire the effect of a previous count becomes more noticeable, due to the formation of extended " ring " systems (*e. g.* Photograph (*d*), Pl. XXVI.) or even continuous " cylinders " of positive ions (*e. g.* Photograph (*c*), Pl. XXVI.). These are undesirable in photographs intended to show interesting particle events in a cloud chamber, and from this point of view the most suitable operating region of the counter for particle detection is the proportional region. By keeping the gas multiplication fairly low any single " rings " present do not seriously obscure any part of the cloud chamber.

§ 4. Discussion of Results.

We now describe a series of photographs which demonstrate the potentialities of the internal counter technique for detecting particles inside a cloud chamber. The internal counter was used in the normal mode of operation, with the threefold telescope arrangement described in § 2 and with the discriminator biased towards events of high total ionization.

The photographs may be divided into three types :

Type (*a*). Photographs showing the internal counter being triggered by single heavily ionizing particles.

Type (*b*). Photographs of local showers and air showers in which two or more fast particles of minimum ionization traverse the internal counter simultaneously.

Type (*c*). Photographs of single fast particles which produce slow knock-on electrons in the gas or in the cathode rods of the counter.

Type (*a*). *Slow Particles.*

Photographs (*e*) and (*f*), Pl. XXVII., show slow particles, probably protons, triggering the internal counter. The slow particle in (*f*) appears to originate in a nuclear star in the glass cylinder of the cloud chamber.

Photograph (*g*), Pl. XXVII., shows an event which is almost certainly the decay of a μ-meson in the glass cylinder at the bottom of the chamber. The decay electron re-enters the chamber, travels upwards at a rather steep angle, unfortunately misses the counter and passes out of the illuminated region of the chamber. Both tracks project back to a point in the cylinder.

Photograph (*h*), Pl. XXVII., shows a slow particle which stops within the sensitive volume of the counter. It is out of the illuminated region at the top of the chamber. From the observed scattering in the gas the particle is very probably a slow meson. No decay electron is visible in the chamber and unfortunately the end of the track is obscured by the large bead of ionization near the wire. For a given quenching delay, the size of the bead round the wire increases with the amount of initial ionization in the counter. As the operating conditions of the counter which affect the size of the bead (in particular the anode voltage and the gas purity) are not necessarily the same in all the photographs, too strict a comparison of bead sizes is not justified. However, this bead is appreciably larger than any obtained under the same conditions. Mutual repulsion of the positive ions in the dense space charge before condensation occurs is more likely to be the cause of the unusually large diameter of the bead, rather than their motion under the influence of the applied field in the few microseconds before it is removed. The spherical nature of the bead supports this view.

In Photograph (*i*), Pl. XXVIII., a slow electron is seen to discharge the counter. This shows that it is possible to trigger a cloud chamber by events which involve electrons or β-particles having energies too low for detection by conventional techniques.

Altogether 26 slow particles (excluding electrons) of ionization greater than about three times minimum have been photographed in a total sensitive time of 727 hours—corresponding to a vertical intensity of about $9 \cdot 3 \times 10^{-5}$ cm.$^{-2}$ min.$^{-1}$ sterad.$^{-1}$, that is, about 0·01 per cent of the total vertical cosmic ray flux at sea level. In three cases there are definite indications that the heavily ionizing particles are mesons (*e. g.* Photographs (*g*) and (*h*), Pl. XXVII.). In the remaining cases, however, the particles cannot be identified with certainty. On account of the material above the internal counter (chamber wall, external counters, etc.), the slow mesons observed would have incident momenta in the range $(9 \cdot 2 – 10 \cdot 1) \times 10^7$ eV./c., while the protons would lie within the range $(4 \cdot 3 – 6 \cdot 0) \times 10^8$ eV./c.

It is difficult to compare the above rate with the results of other workers obtained under different experimental conditions, but Montgomery (1949) in discussing the rather indefinite evidence available has suggested a value of the order 6×10^{-4} cm.$^{-2}$ min.$^{-1}$ for the omnidirectional intensity of slow particles at sea level. Anderson and Neddermeyer (1936) and Widhalm (1940) have shown that the heavily ionizing particles are much less concentrated about the vertical direction than is the total ionizing component, so the unidirectional intensity is probably of the order 10^{-4} particles cm.$^{-2}$ min.$^{-1}$ sterad.$^{-1}$, a value which is in satisfactory agreement with our observed rate.

Rossi, Sands and Sard (1947) have measured the slow meson intensity at sea level in the momentum range $(6-12) \times 10^7$ eV./c. and obtained a value $5 \cdot 4 \times 10^{-5}$ slow mesons cm.$^{-2}$ min.$^{-1}$ gm.$^{-1}$ (range in air). This corresponds to an omnidirectional intensity of $14 \cdot 0 \times 10^{-5}$ cm.$^{-2}$ min.$^{-1}$ in our momentum range. The unidirectional intensity in this momentum range is therefore probably of the order 5×10^{-5} slow mesons cm.$^{-2}$ min.$^{-1}$ sterad.$^{-1}$, since the angular distribution of slow mesons is likely to be such that there is a general concentration about the vertical. (On the other hand, the angular distribution of slow protons may be expected to be much more isotropic, since many are secondary particles emitted from nuclear evaporations in local material (Anderson and Neddermeyer, 1936). Rochester and Bound (1940) measured the number of slow protons in the range $(3 \cdot 9 – 5 \cdot 5) \times 10^8$ eV./c., and from their results Rossi (1948) has deduced a rate of $(3 \pm 1) \times 10^{-5}$ slow protons cm.$^{-2}$ min.$^{-1}$ sterad.$^{-1}$. These results for the intensities of slow mesons and slow protons taken together give a total slow particle rate which is again in fair agreement with the present results for comparable momentum ranges. It seems reasonable to suppose that the numbers of slow mesons and slow protons observed with the present experimental arrangement are of a similar order.

Type (b) Showers.

The internal counter obviously cannot distinguish between one slow particle having an ionization three times minimum (say) and three simultaneous particles of minimum ionization. The threefold telescope arrangement is therefore sensitive to showers, either incident from the air or locally produced. Examples of events of this type are shown in Photographs (*j*) and (*k*) of Pl. XXVIII. It will be noticed that when particles pass simultaneously through the sensitive volume of the counter, each produces a local discharge near the wire. A particle traversing the counter some short time before the cloud chamber was triggered would be associated with a " ring " of positive ions, more or less expanded (*e. g.* Photograph (*f*), Pl. XXVII.), while a particle arriving after the counter voltage was removed would produce no discharge. This ability to determine whether two or more particles are contemporary decreases the effective resolving time of the chamber and could be employed with advantage in certain experiments. A preliminary estimate of the time interval required for the positive ions to move outwards from a " bead " to a recognizable " ring " gives the effective resolving time of the order 100 microseconds, whereas the resolving time of a normal chamber of this size is of the order 1/50 second.

Type (*c*). *Knock-on electrons.*

Photograph (*l*) of Pl. XXVIII. is typical of many photographs obtained with the experimental arrangement described : a fast cosmic ray particle produces a slow knock-on electron in the gas. The ionization of the fast particle alone would not have given a sufficiently large pulse to pass the discriminator, and it is clearly the slow electron which contributed most to the large pulse obtained. Slow knock-on electrons in the gas or cathode rods of the counter occur fairly frequently, and must be responsible for many of the large pulses obtained from the normal enclosed type of proportional counter. (This conclusion was previously reached by Cook (1949) in some recent work using proportional counters). In experiments using conventional proportional counters there is often an undesirable uncertainty in knowing exactly what events triggered the counter. This is avoided in the present technique.

§ 5. THE INTERNAL COUNTER TECHNIQUE FOR DETECTING NUCLEAR DISINTEGRATIONS.

The internal counter need not, of course, be operated in coincidence with an arrangement of external Geiger counters as in the present work. For example the cloud chamber may be triggered by single pulses from the internal counter alone. With the discriminator set to record only events of high ionization a short run on these lines produced photographs mainly of contamination α-particles. This arrangement would undoubtedly be also sensitive to nuclear disintegrations occurring in the gas of the cloud chamber. A rough calculation based on photographic plate evidence (George and Jason 1949) gives the rate of nuclear stars in the gas of the cloud chamber as four per day. A comparison of this with the measured α-particle rate of the counter, gives an expected yield of only one star photograph per 500 photographs of contamination α-particles. (Cohen has reported a similar difficulty with his combination cloud- and ion-chamber). To bias the counter towards events of higher ionization than that of a contamination α-particle would of course mean that many stars would not be recorded, especially those whose origins lay outside the sensitive volume of the counter.

However, an arrangement consisting of two or more open counters operated in coincidence inside a cloud chamber would be sensitive to stars but not to contamination α-particles provided the counters were separated by a distance greater than the range of such α-particles in the gas ($\sim$5 cms.). At the same time there would be no necessity to bias towards events of very high total ionization. It is hoped to employ this arrangement in the near future.

§ 6. Conclusions.

The technique of controlling a Wilson cloud chamber by an internal counter of the type described has the following features :

(*a*) The counter may be used in the proportional region and set to record either single particles of minimum ionization or heavily ionizing events.

(*b*) Unless a particle happens to hit one of the cathode rods the counter has no stopping power beyond that of the gas filling. Slow particles traversing the cloud chamber and events producing slow particles within it can therefore be photographed.

(*c*) The counter may be operated in such a way as to obtain tracks throughout the sensitive volume of the counter, and thus the useful volume of the cloud chamber is not reduced. This is an important feature, especially since it is proposed to put two or more of these counters inside a cloud chamber to detect nuclear disintegrations in the gas.

(*d*) It is possible to observe the exact event which triggered the counter.

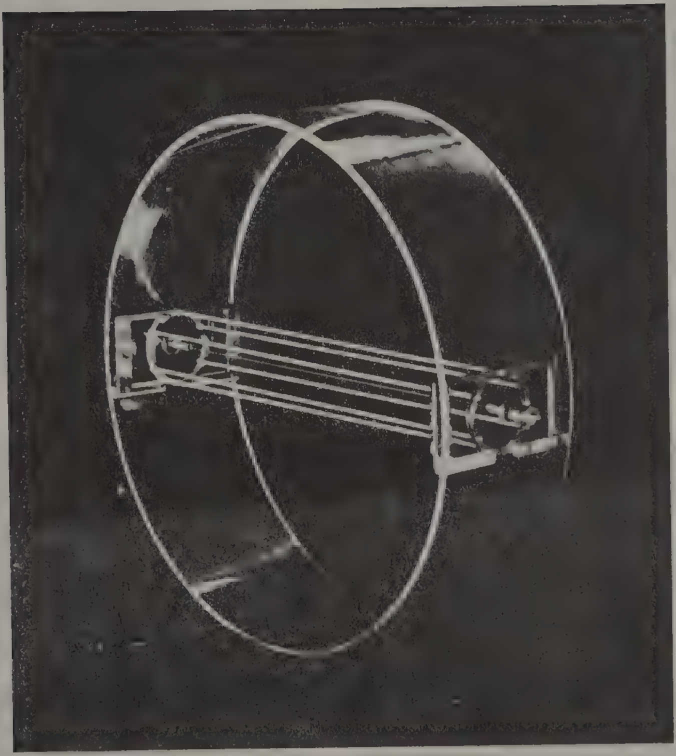
(*e*) The technique is one which is readily applicable to the study of the discharge phenomena in counters.

(*f*) The internal counter affords a means of testing the contemporary nature of particles which pass through it, and the effective resolving time of the cloud chamber is thereby greatly reduced.

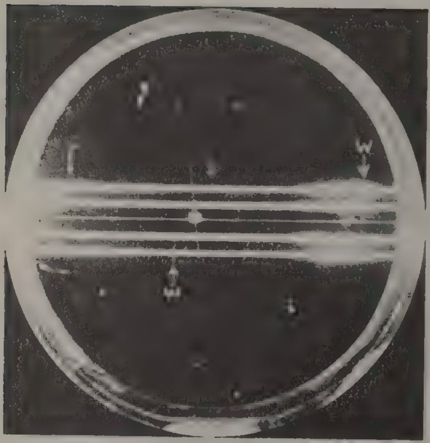
(*g*) The technique is not difficult and, given the few elementary precautions described, the counter is quite stable.

§ 7. Acknowledgments.

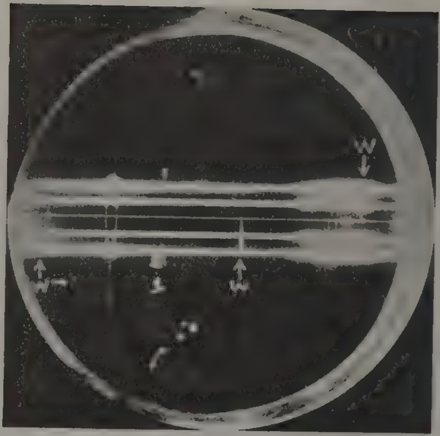
We wish to thank Professor P. M. S. Blackett, F.R.S., for the interest he has taken in this work and for allowing us to make use of the excellent facilities of his laboratory. We are also grateful to Dr. J. G. Wilson for his continued interest and for helpful comments, and to Dr. Nora Page for assistance in some of the initial work. One of us (A. L.) wishes to



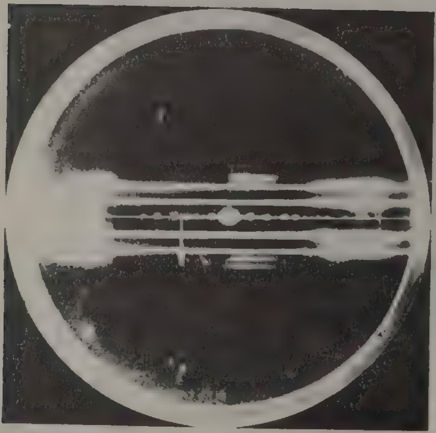
To face page 836



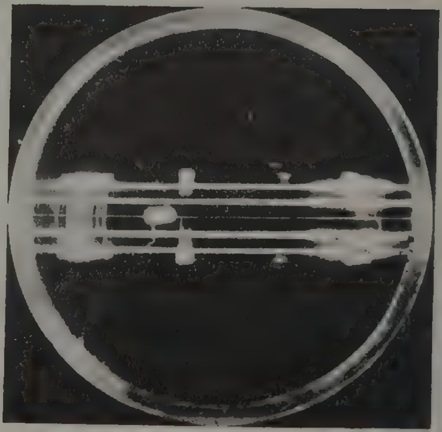
(*a*)



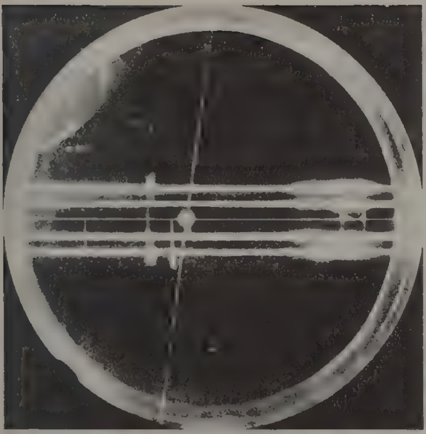
(*b*)



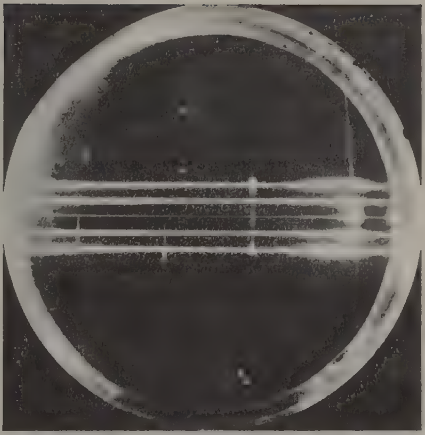
(*c*)



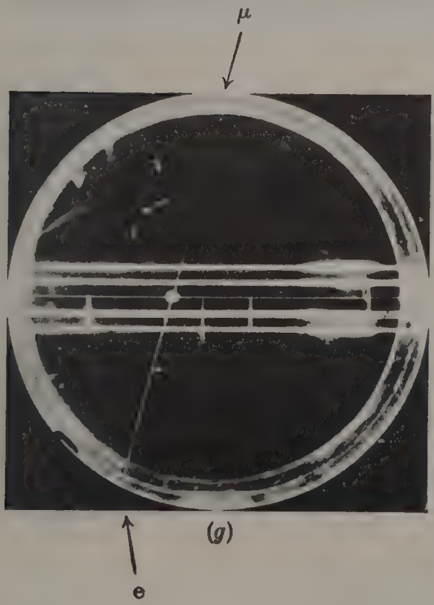
(*d*)



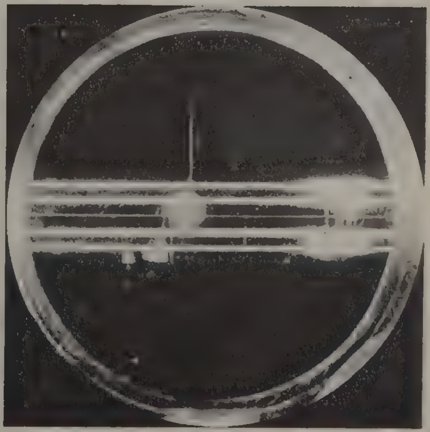
(e)



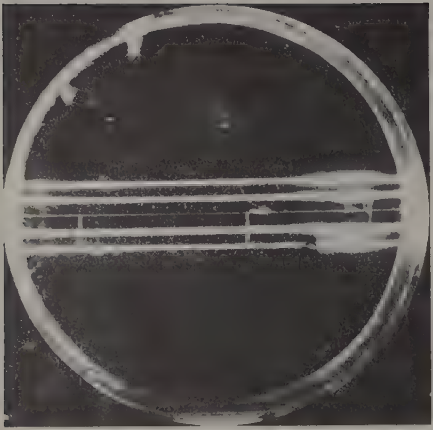
(f)



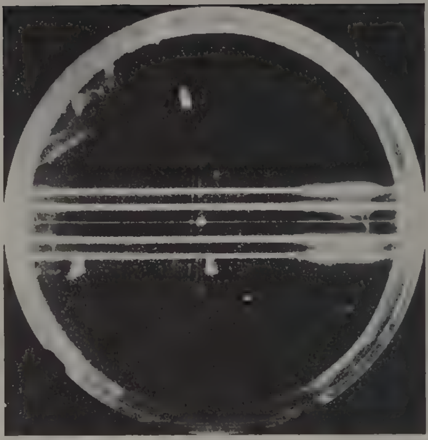
(g)



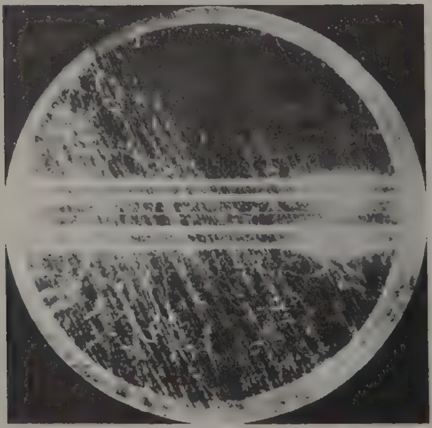
(h)



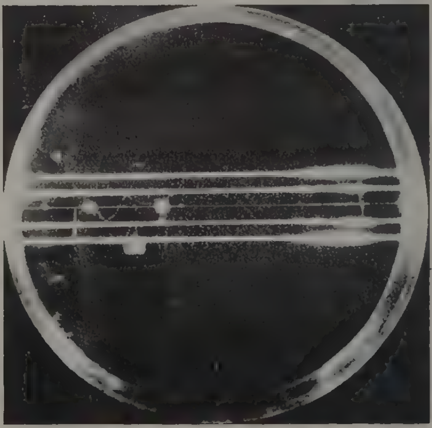
(*i*)



(*j*)



(*k*)



(*l*)

thank the University of Manchester for the award of the John Harling Fellowship, and one of us (N. V. R.) is indebted to the Nuffield Foundation for the award of a Dominion Travelling Fellowship.

REFERENCES.

ADAMS, R. V., ANDERSON, C. D., LLOYD, P. E., and SAXENA, R. C., 1948, *Rev. Mod. Phys.*, **20**, 334.
ANDERSON, C. D., and NEDDERMEYER, S. H., 1936, *Phys. Rev.*, **50**, 263.
BALAKRISHNAN, C., and CRAGGS, J. D., 1950, *Proc. Phys. Soc.* A, **63**, 358.
COHEN, M. J., 1948, *Phys. Rev.*, **74**, 1244; 1949*a*, *Ibid.*, **75**, 1329; 1949*b*, *Biennial Report* (July 1947–June 1949), Palmer Physical Laboratory, Princeton University.
COOK, J. R., 1949, unpublished.
CORSON, D. R., and WILSON, R. R., 1948, *Rev. Sci. Inst.*, **19**, 207.
GEORGE, E. P., and JASON, A. C., 1949, *Proc. Phys. Soc.*, A, **62**, 243.
HODSON, A. L., and LORIA, A., 1949. *Report to the Como Conference on Cosmic Rays.*
McCUSKER, C. B. A., 1944, *J. Sci. Inst.*, **21**, 120.
MONTGOMERY, D. J. X., 1949, *Cosmic Ray Physics* (Princeton: University Press), p. 288.
ROCHESTER, G. D., and BOUND, M., 1940, *Nature, Lond.*, **146**, 745.
ROSSI, B., 1948, *Rev. Mod. Phys.*, **20**, 537.
ROSSI, B., SANDS, M., and SARD, R., 1947, *Phys. Rev.*, **72**, 120.
STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.
WIDHALM, A., 1940, *Zeits. f. Phys.*, **115**, 481.

LXXV. *The Nature of the Neutral Particle Emitted in the Decay of the π-Meson.*

By C. O'Ceallaigh*.

The H.H. Wills Physical Laboratory, University of Bristol†.

[Received July 3, 1950.]

[Plate XXIX.]

Summary.

In order that momentum and energy be conserved in the process of π–μ-decay at rest, it is necessary to assume the emission of a single additional non-ionizing particle in a direction opposite to the initial line of motion of the μ-meson. There exists strong experimental evidence for the conclusion that the neutral particle has small or negligible rest-mass. Hence. we may suppose that it is either (*a*), some light neutral particle possibly identical with the neutrino of β-decay, or, (*b*), a photon.

Some 250 suitable cases of π–μ-decay in Ilford G5 Nuclear Research plates have been examined with a view to testing the hypothesis (*b*). A careful search has been made in the requisite direction, for associated pairs of electrons. In a total scan-length of 38 cms. no such case was observed. Statistical analysis shows that, if indeed a photon were emitted, the probability that it would fail to produce a pair under the conditions of the experiment was 4×10^{-3}.

Since this result would appear to exclude the possibility of photon emission, it is inferred that the π-meson normally decays in accordance with (*a*).

It is shown that the present experiment has much in common with those devised to demonstrate the existence of the neutrino of β-decay.

§ 1. The Decay of π-Mesons.

In the μ-decay of π-mesons as observed in photographic emulsions, it is known that the distribution in range of the μ-particles is consistent with the assumption that they are always emitted with constant kinetic energy, the departures of the individual values of the range from the mean being due to "straggling". It follows from this result that the disintegration is a two-body process, momentum, energy and charge being conserved by the emission of a single neutral particle, which leaves no recognizable trace in its passage through the emulsion.

---

* On leave of absence from University College, Cork.

† Communicated by Professor C. F. Powell, F.R.S.

A study of many examples of the process as it ordinarily occurs in the plates has shown that the spatial distribution of the trajectories of the μ-meson is isotropic with respect to the final direction of motion of the disintegrating π-meson (Lattes *et al.*, 1947, Bradner 1949). Taken in conjunction with the constancy in the range of the μ-meson, this result proves that the π-mesons are commonly moving with very small velocities at the instant of decay. If it were otherwise, the range of the μ-mesons would vary widely according to their direction of emission with respect to the line of motion of the parent π-particles. The neutral particle formed in the decay will therefore be projected in a direction opposite to the initial line of motion of the μ-meson.

MASS OF THE NEUTRAL PARTICLE.

By application of the conservation laws it may be shown that the rest-mass (n_0) of the neutral particle bears a ratio (P) to that of the μ-meson given by

$$\mathrm{P}=n_0/m_\mu=\sqrt{(\mathrm{Q}^2-2\mathrm{QB}_\mu+1)} \quad . \; . \; . \; . \; . \; . \; . \quad (1)$$

where $\mathrm{Q}=m_\pi/m_\mu$, and $\mathrm{B}\mu$ is the value of $(1-\beta^2)^{-1/2}$ proper to the μ-meson at the instant of its formation. The following Table calculated from (1) gives, for certain pairs of values of Q and B_μ, some real values of the mass of the neutral particle expressed in terms of the mass of the electron, m_e.

TABLE I. n_0/m_e.

$\mathrm{B}\mu$ Q	1·036	1·037	1·038	1·039	1·040
1·325	21·2	18·3	14·7	10·0	—
1·321	18·7	15·3	10·8	1·2	—
1·317	15·8	11·6	4·2	—	—
1·315	14·2	9·2	—	—	—
1·313	12·3	6·0	—	—	—
1·312	11·3	3·4	—	—	—
1·311	10·1	—	—	—	—

The most recent measurements (Berkeley Group 1950) give $m_\pi=276\pm6m_e$; $m_\mu=210\pm4m_e$; $\mathrm{Q}=1{\cdot}315$. They are in good agreement with those found for cosmic-ray mesons in this laboratory (Franzinetti 1950). The observed range of the μ-meson in Ilford G5 emulsion is 595 ± 10 microns. Using the established range-energy relationship (Fowler 1950), and setting $m_\mu=210\ m_e$, we find $\mathrm{B}_\mu=1{\cdot}038$.

It will be clear from Table I. that an exact determination of n_0/m_e in the region $n_0 \to 0$ requires very precise knowledge of the values both of Q and of Bμ. Nevertheless, the accuracy of the experimental measurements seems to be such that we may conclude with confidence that the neutral particle has a mass not greater than a few electron masses. Since there is no certain evidence for the existence of neutral particles with a mass of this order, it appears preferable to suppose that it has negligible rest mass. This assumption is consistent with all present experimental information and will be made in what follows.

In order to conserve momentum the neutral particle, if of negligible rest-mass, must have energy=30·1 MeV. Such a particle might be (*a*) the neutrino of β-decay, (*b*) some light neutral particle other than the neutrino or (*c*) a photon.

If it is the neutrino of β-decay, its interaction with matter would appear to be negligible (Nahmias 1935, Ellis and Wooster 1927), and thus it would be expected to escape detection in the photographic emulsion. Evidence based on the shape of the β-ray spectrum of tritium indicates that its rest-mass is probably less than $1/30\ m_e$ (Nielsen 1949).

If the neutral particle is analogous to the neutrino, but not identical with it, it is impossible to state, a priori, in what manner it will interact with matter, save in respect of its energy loss by ionization which will be negligible.

On the other hand, if the particle is a photon, its interaction with matter, though small, will not be negligible. For a photon of 30 MeV., the most important interaction will lead to the creation of electron pairs. Later, it will be shown that the cross-section for the production of a single Compton electron is only 1/7 that for pair production. For this reason we shall be justified in disregarding such of these as may be produced under the conditions of the experiment. It has been the aim of the present investigation to detect pairs of electrons, if they exist, by scanning under high power in a direction opposite to that of the initial path of the μ-meson. The observations would appear to exclude the hypothesis that the neutral radiation emitted in the decay of the μ-meson consists of photons, and strong evidence is thus provided for the existence of some form of "neutrino". That this is identical with the neutrino of β-decay is not certain, but the two particles will be shown to have several properties in common.

§ 2. EXPERIMENTAL METHOD.

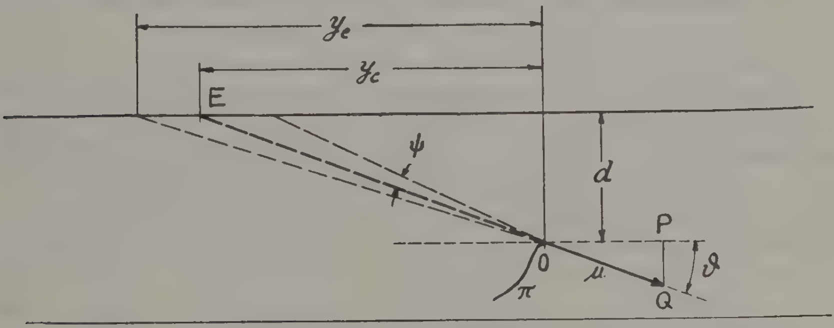
(a) *Geometry of Problem.*

The search for electron pairs associated with π-decay was carried out in the batch of Ilford G5 electron-sensitive plates referred to in previous communications from this laboratory (Camerini *et al.*, 1950). About 400 cases of π-μ-decay were examined. On each of these, measurements were made of the position in the emulsion of the origin of the μ-meson and of its apparent angle of dip using the first 25μ of track. Tracks of which the dip was greater than 15° were then rejected.

The emulsion of these plates was 400μ thick before processing, and after processing, 160μ. It has been shown (Fowler 1950) that for regions of the plates more than 1 cm. from the edges, this shrinkage takes place only in a direction at right angles to the plane of the emulsion and is a uniform linear transformation in this coordinate. We define the shrinkage factor $t_u/t_p=S=400/160$ for the plates concerned. Here, and subsequently, the subscripts u and p will refer respectively to the unprocessed and processed emulsions.

Fig. 1 represents a vertical section of the emulsion. The path of the neutral particle through the unprocessed emulsion will be OE$=t$, O being the origin of the μ-meson, and E the point at which the particle leaves the emulsion. The angle of dip, $\vartheta=\text{arc tan PQ}/25$, PQ being measured in microns. We have the following relationships valid for processed and unprocessed emulsions:

Fig. 1.



$$\tan\vartheta_u=S\tan\vartheta_p \qquad (2)$$

$$y=d_u\cot\vartheta_u=d_p\cot\vartheta_p \qquad (3)$$

$$t_u=y(1+\tan^2\vartheta_u)^{1/2} \qquad (4)$$

In the actual experiment the most convenient distance to measure is y. For each track, ten measurements of dip were made by means of an eyepiece scale, and the graduated focusing drum of the focusing scale of the instrument. The distance d_p was measured for the same setting of the instrument. From these data, the value of y was computed using equation (3). The path-length thus estimated is less than the actual path t_u in the unprocessed emulsion, but it may be seen from (4) that the defect of y from t_u is small for the values of ϑ included in the measurements, and further, it may be allowed for.

(b) *Method of scanning and allowance for multiple scattering.*

The scanning of the plate was carried out under magnifying power objective × 45, with binocular eyepieces × 8. A hair-line in one eyepiece was set so that its direction was parallel to the y-axis of motion of the microscope stage. The plate was then rotated until the hair-line appeared to lie evenly along the mean-line of the grains in the first 25μ of the μ-meson track. The field of view extended to 100μ on either side of the hair-line.

The path followed by the neutral particle will be tangent to the μ-meson track at the point of origin. In general, because of multiple scattering, the directions of the mean-line and of the tangent will not coincide. The angle between these directions will be normally distributed about the direction of the mean-line and the variance of the distribution may be computed from the momentum of the scattered particle and the composition of the emulsion (Rossi and Greisen 1941). Calculation shows negligible probability that the projection of the tangent lies outside a sector of angle 2° on either side of the mean-line, with centre on the origin of the μ-meson. For values of $y \leqslant 2500\mu$ it may be seen that the whole of this sector lies within the 200μ strip defined by the field of view. For longer scans, 1 cm. $> y \geqslant 0{\cdot}25$ cm., two additional $200\,\mu$ strips were scanned on either side of the central strip.

The estimate of y is, in similar manner, uncertain because of the effect of multiple scattering. Referring to fig. 1 it will be clear that this estimate, say y_c, depends on a measurement of the dip of the projection on a vertical plane, of the *chord* of the first $25\,\mu$ of μ-meson track. A correct estimate, say y_t, would be based on the dip of the *tangent* at the origin. It may be shown that the angle between chord and tangent has a normal distribution about the direction of the chord, and has variance $\langle\psi^2\rangle = \frac{1}{3}\langle\phi^2\rangle$ (Rossi and Greisen 1941). $\langle\phi^2\rangle$ is the mean square value of the projection of the angle between the tangents at the end-points of the segment. The value of $(\langle\psi^2\rangle)^{1/2}$ for the conditions of the experiment was calculated to be 1·8°.

For any individual track, y_t may be greater or less than the observable y_c. For tracks of sensible dip, because of the small half-width of the distribution of ψ, this difference is small or negligible. It is otherwise for tracks of small dip, and especially for "flat" tracks ($\vartheta_p \simeq 0$), and must be taken into account. This was done by finding the distribution function for y in terms of d, and ϑ (fig. 1) assuming normal distribution of ψ. The mean value $\langle y\rangle$, of y was obtained for this by graphical integration. For tracks of measurable dip the scan was continued to $y_e > y_c$ (fig. 1) such that the mean value of y, computed as above, was equal to y_c. For "flat" tracks, which by their nature make large contribution to the length scanned, it was found possible to express $\langle y\rangle$ in terms of a closed expression involving the tabulated functions $\mathrm{E}rf(x)$ and $\mathrm{E}i(x)$.

§ 3. EXPERIMENTAL RESULTS.

The total number of π-μ decay events selected as suitable was 253. The sum of all the effective scan-lengths Σy_c computed as above was 36 cms. The corresponding true path-length in the emulsion Σt_u was 38 cms. 35 pairs of particles, probably electron pairs, were observed to have their points of origin within the scan strip-width (but not necessarily within the sector). The total area of the scan was

1·32 cm². This yields a density of 26 pairs per cm². of these plates. Other workers in this laboratory (Carlson *et al.*, 1950) report a density of 24 pairs/cm². The agreement is satisfactory. Neither figure has been corrected for failure to observe the small proportion of very steeply dipping pairs which are bound to escape detection.

For 24 cms. of this scan length each field was scanned twice and for the remaining 12 cms. the field was scanned once only. The corresponding areas of scan were calculated to be, respectively, 1·0 cm². and 0·32 cm². The number of pairs observed in the first series was 28 and, in the second series, 7. It may be concluded from these figures that there was no significant reduction in the efficiency of detecting pairs introduced by the change to single scanning. The apparent disproportion between scan-length and scan-area arises from the facts that the ratio scan-area/scan-length increases with the cube of the scan-length, and that the incidence of long scans was slightly smaller in the second series.

Theoretical Cross-Sections.

It has been shown (Heitler 1944) that, for high energies, both electrons of a pair are emitted in the forward direction. The solid angle Ω into which the electrons are emitted is of the order $\Omega \sim m_e c^2/\mathrm{T}$, where T is the total energy of the pair. For 30 MeV. photons $\Omega \sim 1/60$.

The cross-section for pair-production σ_{pair} by 30 MeV. photons $= 7\overline{\mathrm{Z}}^2 \times 5{\cdot}71 \times 10^{-28}$ cm², and for Compton Effect

$$\sigma_{\text{Comp}} = 3 \times 10^{-2} \overline{\mathrm{Z}} \times 6{\cdot}57 \times 10^{-25} \text{ cm}^2.$$

From data supplied by the manufacturers of the G5 plates we may calculate $\overline{\mathrm{Z}}^2 = 451$, $\overline{\mathrm{Z}} = 13{\cdot}1$. The number of atoms per c.c. of emulsion $= 8{\cdot}37 \times 10^{22}$. We obtain from this $\sigma_{\text{pair}}/\sigma_{\text{Comp}} = 7$, and a probability of pair production per cm. path in the emulsion$=0{\cdot}152$. Hence, if the neutral particle is a photon, we should expect to observe, on the average, 5·77 electron pairs " associated " with the cases of π–μ-decay examined.

The criteria of " association " are somewhat complicated by the multiple scattering of the μ-meson, but for present purposes, may be formulated with sufficient accuracy as follows :

(1) The origin of such a pair must be found within a solid angle containing OE (fig. 1) of such magnitude as to allow, in any individual case, for uncertainty as to the direction of the tangent at the origin of the μ-meson track. The projection on the surface of the emulsion of this solid angle is the sector already referred to.

(2) The pair must lie within a small calculable solid angle with OE as axis since the electrons are projected in the forward direction.

(3) It must be a pair of electrons of energy 30 MeV. to within the experimental error of measurement.

36 pairs were examined in respect of these criteria. All, save one, were rejected because of violation of (1) and (2). One pair remains which satisfied (1) and (2) but did not satisfy (3). This event will be fully

described in the appendix to this paper where it will be shown that it is most unlikely to be a pair of electrons. It has not been found possible to furnish a precise identification of its nature.

It is extremely improbable that any "associated" pairs escaped detection. Firstly, such pairs, if produced, would be more favourably situated for observation than the background pairs. Their dip will be small on account of the exclusion of steeply dipping μ-mesons. Furthermore, even if their points of origin were undetected in one field, their direction would be such that they would be readily observed as converging pairs of tracks in the succeeding two or three fields. Secondly, as has already been stated, the superficial density of background pairs is in satisfactory agreement with that found by other workers using plates chosen from the same series.

There remains to be considered a statistical possibility namely that the neutral particles are indeed photons, but that they failed to be materialized in a scan-distance in which 5–6 pairs were to be expected. Each scan-length y_j microns corresponding to the jth π–μ event, constitutes a "trial" in which the probability of producing a pair is $1{\cdot}52\times10^{-5}y_j$. The probability of failure in this trial is $(1-1{\cdot}5\times10^{-5}y_j)$ and the probability of failure in all n trials $\mathrm{P_F}=\prod_{j=1}^{n}(1-1{\cdot}5\times10^{-5}y_j)$. A sufficient approximation to $\mathrm{P_F}$ will be Poisson's Function of Rare Statistical Frequency applicable to the case where all trials have equal probability. This yields $\mathrm{P_F}=\epsilon^{\ \mu}=\epsilon^{-5\cdot77}=4\times10^{-3}$. Rejecting this supposition on the grounds of its extreme improbability, we may conclude with reasonable confidence that the neutral particle, whatever be its nature, is not a photon.

§ 4. INTERPRETATION OF RESULTS.

(a) *Spin of the π and μ-mesons.*

Further support for the reasonableness of this conclusion appears from considerations of spin conservation. Experiment has shown that the μ-meson decays into an electron and more than one neutral particle. (Leighton *et al.* 1949, Davies *et al.* 1949). Evidence based on the average energy of the electron spectrum favours the emission of two neutral particles, presumably identical. The results of these workers would appear to be confirmed by a more extended series of measurements on which the writer is at present engaged. Counter experiments (Hincks and Pontecorvo, 1949, 1950; Piccioni, 1949) appear to exclude the possibility that these are photons. Taken in conjunction with the results of the present work, these findings indicate that the decay chain π-μ-e takes place as follows:

$$\pi\rightarrow\mu+\nu_1$$

$$\mu\rightarrow e+2\nu_2.$$

Whether the spin of ν_2 is integral of half-integral, it follows that that of the μ-meson is half-integral. This result is consistent with evidence based on the distribution curve of burst frequency (Christy and Kusaka, 1941).

If ν_1 has spin integral or zero, it follows that the π-meson has half-integral spin. This is in contradiction to the role of the π-meson in the theory of nuclear force fields which envisages interactions of the type $p \rightarrow n+\pi^+$; $n \rightarrow p+\pi^-$, an assumption supported by experimental evidence for the creation of π-mesons by nucleon–nucleon collisions*. We infer that ν_1 is a fermion (a particle of spin $\frac{1}{2}$), a conclusion not contradicted by the results of the present experiment.

Similarly, we may conclude that if the competing decay process $\pi \rightarrow e+\nu_3$ exists, it is likely to lead to the production of monoergic electrons unless an odd number >1 of neutral particles is emitted. The theoretical cross-section for this process is of the order 1/100 that for the chain π-μ-e (Ruderman and Finkelstein 1949).

(b) *Existence of the Neutrino.*

It appears that the characteristics of the decay of π-mesons may have a close bearing on this problem. It would be tempting to identify, ν_1, ν_2 and ν_3 with the neutrino of β-decay. In common with this particle, it would seem that they are fermions of negligible rest-mass. However, the present state of theory is such that perhaps it would be unwise to press the analogy between π-decay and the β-decay of radioactive elements.

It is a necessary consequence of the assumed properties of the " neutrino ", that such particles have a very weak interaction with matter. Attempts have therefore been made to prove the existence of the particles indirectly. It has commonly been considered that satisfactory evidence would consist in a demonstration that the conservation of momentum, as well as that of energy, requires the assumption of the emission of a neutral particle in the β-decay of a nucleus, or in the process of K-capture.

Several such experiments have been described involving the β-ray spectra of Cl^{38} (Crane and Halpern 1938, 1939) Kr^{88} (Jacobsen and Kofoed-Hansen, 1945) and on the formation of Li^7 by K-capture from Be^7 (Allen 1942), the magnitude of the recoil momentum has been shown to be consistent with the neutrino hypothesis. These experiments, though brilliantly conceived, are attended by great difficulties, because the disparity between the masses of the electron and the nucleus has the consequence that the energy of recoil of the latter is very small.

The present experiment has points of similarity with that of Allen in which the recoil momentum of the Li^7 was shown to be due entirely to the emission of the neutrino. The recoil energy was only about 50 eV. and was measured electrostatically. The present method is similar in principle, but the μ-recoil has energy 4·1 MeV., is visible in the emulsion and its momentum is measured by use of the well established range–energy relationship. If subsequent developments should establish the fact that of the neutral particle is identical with the β-decay neutrino, the present experiment could be regarded as supplementing and confirming the β-ray experiments. In view, however, of the anomalous nature of the particle apparently associated with a case of π-μ-e decay, already referred to, and described in detail in the Appendix, it would seem at present unwarranted to conclude that the neutral particles emitted in meson decay are to be identified with the "neutrino" postulated by Fermi and by Pauli in order to explain the observed phenomena of β-decay.

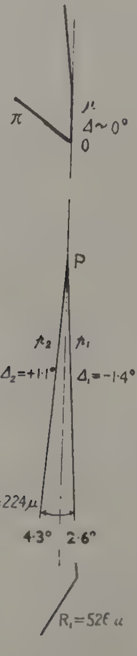
---

* *Note added in proof*—Additional direct experimental evidence for the conclusion that the π-meson is not a fermion comes from the recent work of Panofsky, Aanodt and York (*Phys. Rev.*, **78,** 825, 1950) on the γ-radiation associated with the absorption of π^--mesons in hydrogen under high pressure. This would appear to demonstrate the occurrence of the reaction $\pi^-+p \rightarrow n+\gamma$.

APPENDIX.

In the course of the search for electron pairs which could be attributed to γ-rays liberated by the decay of π-mesons, an event was observed of which photo-micrographs are reproduced in Pl. XXIX. Although no definite conclusion can be drawn from it, it displays a number of remarkable features, and it was therefore found desirable to describe it in detail :

Fig. 2.



1500 μ from the point of arrest, O, of a π-particle in the emulsion, and in the line of motion of the μ-meson produced by the decay of the π-particle, and which subsequently decayed itself into an electron of 36 MeV., a pair of particles p_1 and p_2 appear to originate at a point P. OP is indistinguishable from the initial line of motion of the μ-meson. Further, OP is inclined to the plane containing the initial directions of motion of the two particles p_1 and p_2 at an angle less than 1°. These relationships, the angles of dip (Δ), positive if towards glass, of tracks p_1 and p_2 and the lengths (R) of the tracks are shown in fig. 2.

Although the ranges of the two particles in the emulsion are relatively short, it has been found possible to make a determination of the scattering parameter $\bar{\alpha}$ (Fowler 1950), for the tracks are heavily scattered and short " cell-lengths " of 20 μ can therefore be employed. The values of $\bar{\alpha}$ thus obtained are :—for p_1, $\bar{\alpha}_{100\mu}=1{\cdot}63\pm0{\cdot}25$; and for p_2, $\bar{\alpha}_{10c\mu}=1{\cdot}6\pm0{\cdot}35$. The grain-density in the two tracks are $16{\cdot}3\pm3{\cdot}0$ and 20 ± 4 grains per 50 μ, respectively, values which may be compared with those due to particles with minimum ionization recorded in the same plates ; namely, 11·5 grains per 50 μ. The uncertainties quoted are standard deviations.

If the tracks are due to particles with the elementary electronic charge, the observed values of the scattering parameter correspond to particles of kinetic energy ~ 10 MeV., and in view of the considerable limits of error due to statistical fluctuations, it would therefore be reasonable, from the point of view of the kinetic energy of the particles, to attribute them to electrons produced by a 30 MeV. γ-ray. Other features of the observations, however, make it very difficult to accept this explanation.

Firstly, the grain-density in the tracks is exceptionally high if they are due to particles moving at relativistic velocities, as would be the case, over most of the trajectory, if the particles were electrons. Here again, the statistical fluctuations, due to the relatively restricted lengths of the trajectories is rather large, and the evidence is not decisive. Observations, however, on more than 1000 tracks, of length $>3000\,\mu$, of which the grain density has the minimum value, have been made in this laboratory in connection with other investigations. These observations were analyzed to determine the frequency with which the grain-density in the tracks of particles of high energy, measured over a length of 500 μ, gave an apparent grain-density of 16 or greater. In no case was a value observed as high as the mean for the two tracks under discussion.

A second feature of the event is that the two trajectories both appear to terminate in the emulsion without any indication of an increase in the specific ionization due to a gradual reduction in velocity. Such an effect might be explained as due to the annihilation of a positive electron, and to a concomitant large angle deviation in the track of an electron. Since both of the particles cannot be positive electrons, the apparent endings of the tracks have been scrutinized. They occur in the middle of the emulsion where there are no abnormalities of any obvious kind.

A third feature of the observations is the angle between the pair, viz. 9·5°. This angle is very large if the pair is attributed to the materialization of a 30 MeV. γ-ray, but not impossibly so. Because of the occurrence of these three unusual features in the same event, it is reasonable to reject the view that it is due to a pair of electrons. In this case, we have to account for a phenomenon in the line of motion of a neutral particle emitted in the decay of a π-meson with exceptional features not previously observed.

The values of grain-density and scattering constant $\bar{\alpha}$ correspond to masses of the order 30 m_e for both tracks, but, as is well known, statistical fluctuations may lead to serious error in an estimate of mass based on observations on a single relatively short track.

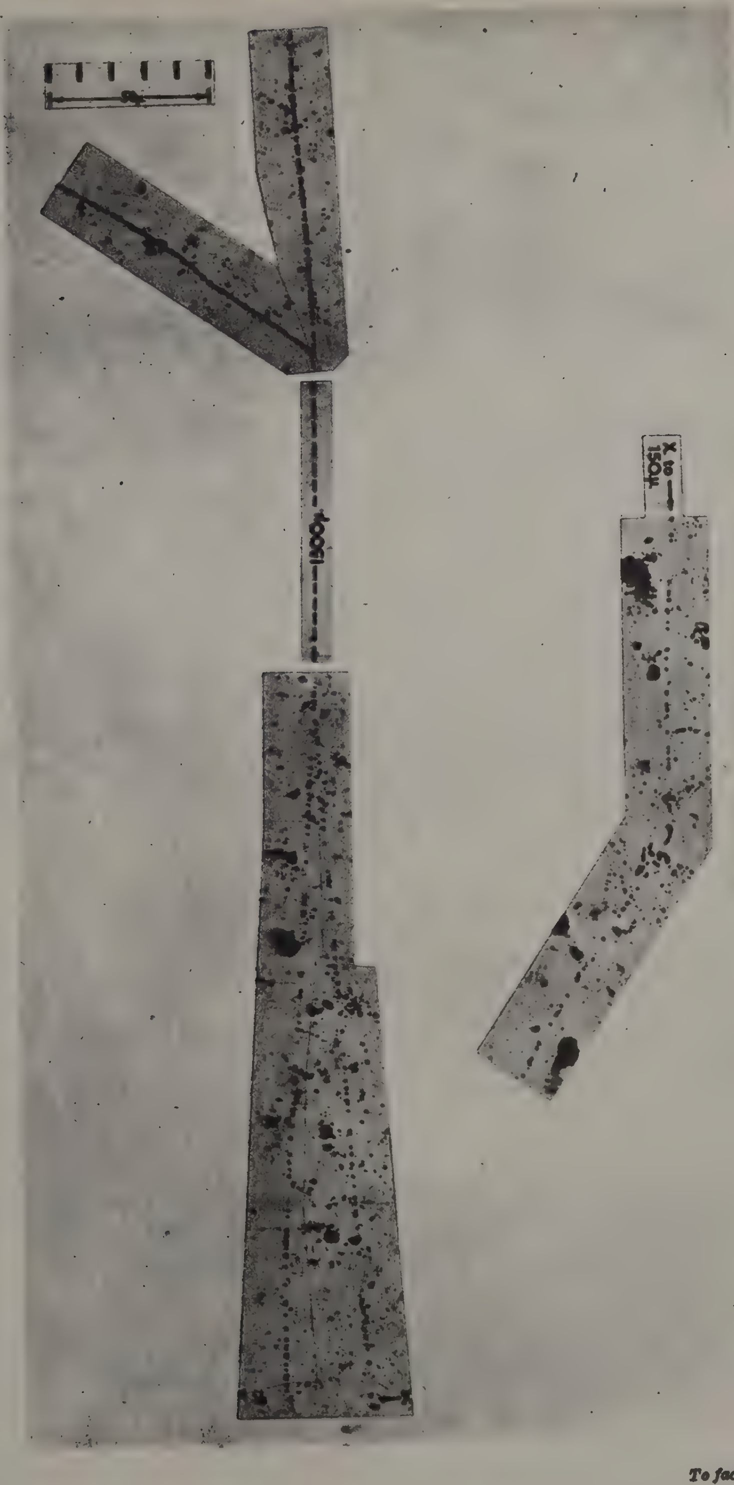
If we accept the evidence of the grain-density and scattering of the track as not being seriously in error, it appears to be very difficult to account for the phenomena in terms of charged particles of which the existence is established. It is also difficult to find a consistent explanation of the observations in terms of known particles which have, by chance, passed through the emulsion during development. Two protons, for example, originating in a nuclear explosion during development would produce thin tracks but they should show evidence of an increase in grain-density towards the end of their range, a feature which is not observed. It appears best, however, to defer an attempt to identify the event until a second observation of the same kind allows us to exclude the possibility that it is to be attributed to a rare but trivial occurrence.

ACKNOWLEDGMENTS.

The writer wishes to thank Professor C. F. Powell, F.R.S. for suggesting the experiment, for extending to him the hospitality of his laboratory and for his continued interest and encouragement. It is a pleasure to record his thanks to his wife who assisted in the experiment and is responsible for most of the dip measurements, to Mr. R. J. M. Gold who gave assistance with the graphical integrations, to the various members of the laboratory research group, particularly Mr. U. Camerini and Mr. J. H. Davies for many helpful discussions, to Miss P. Dyer and Mr. R. Gattiker for preparing the Plate, and, finally to the team of microscope observers. He also wishes to thank the Governing Body of University College, Cork, for leave of absence and a grant towards maintenance.

REFERENCES.

ALLEN, 1942, *Phys. Rev.*, **61**, 692.
BRADNER, 1949, *U.C.R.L.*, 486.
CAMERINI, FOWLER, LOCK and MUIRHEAD, 1950, *Phil. Mag.*, **41**, 413.
CARLSON, HOOPER and KING, 1950, Private Communication
CHRISTY and KUSAKA, 1941, *Phys. Rev.*, **59**, 414.
CRANE and HALPERN, 1938, *Phys. Rev.*, **53**, 789.
DAVIES, LOCK and MUIRHEAD, 1949, *Phil Mag.*, **40**, 1250.
ELLIS and WOOSTER, 1927, *Proc. Roy. Soc.* A, **117**, 109.
FOWLER, 1950, *Phil. Mag.*, **41**, 169.
FRANZINETTI, 1950, *Phil. Mag.*, **41**, 86.
HEITLER, W., *Quantum Theory of Radiation.* Oxford, 1944.
HINCKS and PONTECORVO, 1948, *Phys. Rev.*, **75**, 257; 1949, *Phys. Rev.*, **75**, 698; 1950, *Can. Journ. Res.*, **28**, 29.
JACOBSEN and KOFOED-HANSEN, 1935, *K.D.V.S.*, **23**, No. 12.
LATTES, OCCHIALINI and POWELL, 1947, *Nature*, **160**, 453.
LEIGHTON, ANDERSON and SERIFF, 1949, *Phys. Rev.*, **75**, 1432.
NAHMIAS, 1935, *Proc. Camb. Phil. Soc.*, **31**, 99.
NIELSEN, C. E., 1949, quoted Gamow and Critchfield, *Atomic Nuclei.* Oxford, 1949: Appendix III.
PICCIONI, 1948, *Phys. Rev.*, **74**, 1754.
ROSSI and GREISEN, 1941, *Rev. Mod. Phys.*, **13**, 240.
RUDERMAN and FINKELSTEIN, 1949, *Phys. Rev.*, **76**, 1458.
SARD and ALTHAUS, 1948, *Phys. Rev.*, **73**, 1251; **74**, 1364.
SMITH, BARKAS, BISHOP, BRADNER and GARDNER, 1950, Berkeley group, *Phys. Rev.*, **78**, 86.


50μ
100μ
X 10
150μ

LXXVI. CORRESPONDENCE

An Estimate for the Position of the Lowest Dipole-level of a Nucleus.

By B. F. Touschek,
Department of Natural Philosophy, The University, Glasgow *.

[Received June 20, 1950.]
* Communicated by the Author.

The completeness property of the set of nuclear eigenfunctions can be used to obtain an inequality for the position of the lowest dipole level of a nucleus, *i.e.* for the lowest level that can be reached from the groundstate by a dipole transition.

Feenberg (1935) has applied sum rules to obtain a lower limit for the energy of a hypothetical excited state of He^4. His method can be adapted to apply to any nuclear system, though the actual evaluation of the limiting value for the excitation energy can only be carried out on the basis of a strongly simplified nuclear model. By considering two extreme cases it will be shown that the limiting value of the excitation energy does not depend critically on the model used for its derivation. In this paper the method will be outlined for dipole transitions though, of course, a similar procedure may be applied to quadripole and higher transitions. The estimates obtained for quadripole transitions give a much higher limiting energy than in the dipole-case ($140A^{-2/3}$ MeV. as compared with $50A^{-2/3}$ MeV. for a dipole) and are therefore probably less useful.

The basic idea is to evaluate the sums

$$X = \sum_b |x_{ab}|^2 \quad \quad (1)$$

and

$$H = \sum_b (E_b - E_a) |x_{ab}|^2, \quad \quad (2)$$

by applying the completeness of the set of nuclear eigenfunctions, x_{ab} is the electric dipole moment (after the separation of the motion of the centre of gravity) of the transition $a \rightarrow b$ between the states a and b with energies E_a and E_b of a nucleus. If $a=0$ denotes the groundstate and $b=1$ the first excited state with $x_{01} \neq 0$, then all the $E_b - E_0$ are positive and one has

$$E_1 - E_0 < H/X. \quad \quad (3)$$

In a one particle picture in which the nucleus consists of independent neutrons and protons the sums (1) and (2) can be readily evaluated:

$$X = \frac{R^2 Z}{5} \tfrac{1}{2}(1+I) \quad \text{and} \quad H = \frac{\hbar^2}{2M} Z \tfrac{1}{2}(1+I),$$

where $I = 1 - 2Z/A$ is the isotopic ratio, M is the nucleon mass and R the radius of the nucleus. The factor 1/5 corresponds to the assumption that the nucleons are uniformly distributed over a sphere of radius R. The factor $\frac{1}{2}(1+I)$ takes account of the motion of the centre of gravity: it vanishes for a system consisting of particles of one and the same specific charge. Taking $RA^{-1/3}=1{\cdot}5\times 10^{-13}$ cm. one obtains the estimate

$$E_1-E_0<\frac{50}{A^{2/3}}\,[\text{MeV.}], \quad \ldots\ldots\ldots \quad (4)$$

for the position of the first dipole level of a nucleus.

It is well known that the independent particle model gives a poor description of the sum-over-cross-sections for nuclear photo effects and recent calculations by Levinger and Bethe (1950) on the yield of photo neutrons have shown a strong dependence of the results on the assumed exchange character of nuclear forces. This can readily be understood if one considers the tendency of group formation inherent in a force-picture with saturation properties. This group-formation will result in strong correlations between the coordinates of different particles (in particular between two neutrons and two protons forming an intranuclear α-particle), so that considerable changes may be expected in the values of H and X.

To see how the estimate (4) is affected by different assumptions about nuclear forces one can consider an independent particle model in which the nucleus is assumed to consist of αZ free protons, $\frac{1}{2}Z(1-\alpha)$ α-particles and $A-Z-Z(1-\alpha)$ free neutrons. The parameter α can then be interpreted as the degree of dissociation of the α-particles in the nucleus. A model of this type with a reasonable value of α (Wergelands (1941) results on Be^6 suggest a value $\alpha\sim 0{\cdot}25$) should be expected to give a fair rendering of some of the effects of exchange forces: the formation of α-particles as relatively stable sub-units and the resultant correlations between the coordinates of the nucleons. In this model the sums (1) and (2) can again be evaluated in an elementary manner:

$$\left.\begin{aligned} X &= \frac{R^2Z}{10}\{I+\alpha+3I^2(1-\alpha)\} \\ H &= \frac{\hbar^2 Z}{4M}\{I+\alpha\} \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (5)$$

In the case of complete dissociation of the α-particles this gives, of course, the result already obtained for the free neutron-proton model. The case $\alpha=0$ gives $X=0$ and $H=0$ if $A=2Z$. In this case all the particles must have the same specific charge so that both the cross-section and the average energy-loss of dipole transitions must vanish. From (5) one obtains the refined estimate for the position of the lowest dipole level:

$$E_1-E_0<\frac{5\hbar^2}{2MR^2}\,\frac{1}{1+3I^2(1-\alpha)/(I+\alpha)}, \quad \ldots\ldots \quad (6)$$

and it is immediately obvious that for moderate isotopic ratios this expression is nearly independent of α. Even for very heavy nuclei the difference between the values of (6) for $\alpha=0$ and $\alpha=1$ respectively is not more than 40 per cent. This small variation seems to indicate that the estimate of the position of the first dipole level is better than the theory from which it has been derived and suggests to try to determine a similar limit for quadripole levels by means of the free neutron-proton model. Again an elementary calculation similar to that leading to the inequality (4) gives

$$(E_1-E_0)_{E.Q}<\frac{140}{A^{2\,3}}\,\mathrm{MeV}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

This limit is not very useful for light nuclei since there it corresponds to an energy far in excess of the "ionization-energy" of a neutron. For A$>$70 it should fall into the discrete part ($<$8 MeV.) of the nuclear spectrum so that one concludes that every nucleus with A$>$70 should have at least one quadripole level in the discrete spectrum. The corresponding consideration for dipole levels gives a critical atomic number of A$\sim$16 above which every nucleus should have at least one discrete dipole level. Feenberg's calculations seem to indicate that the actual level usually occurs very near to its theoretical limit. O^{16} with a dipole level at 7 MeV. and a limiting energy of approximately 8 MeV. is an example for this behaviour.

References.

Feenberg, E., 1935, *Phys. Rev.*, **49**, 328.
Levinger, J. S., and Bethe, H. A., 1950, *Phys. Rev.*, **78**, 115.
Wergeland, H., 1941, *Norske Vid. Selsk. Skr.*, 1.

Examples on the Calculation of Energy States by the Uncertainty Relations.

By Hsin P. Soh,
National Chekiang University *.

[Received June 14, 1950.]

Although the Uncertainty Principle forms one of the most fundamental concepts in quantum mechanics, the possibility of making quantitative predictions based on this principle is little explored. In this letter we shall give some examples illustrating the possibilities in this direction.

I. *Hydrogen atom.*

The energy of the system satisfies the inequality

$$E\geqslant\frac{1}{2m}P_r^2-\frac{ze^2}{r},$$

since the contribution of angular motion is not included in the right side of the above expression. The minimum energy is given by the "spread" (mean square deviation) of P_r and that of r, *i. e.*

$$E_{\min}=\frac{1}{2m}\,\varDelta P_r^2-\frac{ze^2}{\varDelta r},$$

---

* Communicated by Professor M. Born, F.R.S.

where ΔP_r is the minimum measurable momentum and Δr is the minimum measurable distance of the electron from any other particle, here the nucleus. The minimum of P_r is now postulated to be connected with that of r by the relation $\Delta P_r \Delta r \geqslant \hbar$ *. Hence

$$E_{\text{min}} = \frac{\hbar^2}{2m}\frac{1}{\Delta r^2} - \frac{ze^2}{\Delta r}.$$

The last expression can be minimized by finding a suitable value of Δr, which may be varied by changing, for instance, the wavelength of the light used in the measurement. Using this minimum value, $\Delta r = \hbar^2/ze^2 m$, we get finally, $E \geqslant -z^2e^4m/2\hbar^2$.

II. *Relativistic hydrogen atom.*

For the relativistic hydrogen atom one has

$$E \geqslant [m_0^2c^4 + c^2P_r^2]^{\frac{1}{2}} - \frac{ze^2}{r} - m_0c^2.$$

The argument in the case of this problem is the same as in the last example, *i. e.*

$$E \geqslant [m_0^2c^4 + c^2\Delta P_r^2]^{\frac{1}{2}} - \frac{ze^2}{\Delta r} - m_0c^2 \geqslant \left[m_0^2c^4 + c^2\frac{\hbar^2}{(\Delta r)^2}\right]^{\frac{1}{2}} - \frac{ze^2}{\Delta r} - m_0c^2.$$

Varying Δr and substituting in this formula, we obtain

$$E \geqslant m_0c^2[(1-\alpha^2z^2)^{\frac{1}{2}} - 1], \qquad \alpha = \frac{e^2}{\hbar c}.$$

III. *Hydrogen atom including the motion of the nucleus.*

In this case

$$E \geqslant \frac{1}{2M}P_R^2 + \frac{1}{2m}P_r^2 - \frac{ze^2}{r_{12}}.$$

The minimum energy is given by the spread of P_R, P_r and r_{12}. But the minimum of r_{12}, the distance of the electron from the nucleus, cannot be smaller than either of the minimum measurable distance of the electron from any other particle, Δr, or of the minimum measurable distance of the nucleus from any other particle, ΔR. Hence if $\Delta r \geqslant \Delta R$, then

$$E \geqslant \frac{\hbar^2}{2M}\frac{1}{\Delta R^2} + \frac{\hbar^2}{2m}\frac{1}{\Delta r^2} - \frac{ze^2}{\Delta r}.$$

---

* In the case of two dimensional motion this expression can be rigorously derived from the wave property of the electron. We shall, however, treat it as a postulate. In dealing with more complicated atoms we shall also postulate $\Delta P_r \Delta r \geqslant n\hbar$ for " higher quantum states "

Now the first term is independent of the other terms and has its minimum value when ΔR has the greatest value, which by our assumption is at most equal to Δr. Hence

$$E \geqslant \frac{\hbar^2}{\Delta r^2}\left(\frac{1}{2M}+\frac{1}{2m}\right)-\frac{ze^2}{\Delta r},$$

and by simple variation, we get

$$E \geqslant -\frac{Mm}{M+m}\frac{z^2e_2}{2\hbar^2}.$$

The same lower limit is obtained if $\Delta r \leqslant \Delta R$ is assumed.

In the three examples given above the lower limit of energy obtained from uncertainty relation agrees in every case with that obtained by other methods.

IV. *Helium-like atoms.*

For helium-like atoms

$$E \geqslant \frac{1}{2m}(P_{r_1}^2+P_{r_2}^2)-ze^2\left(\frac{1}{r_1}+\frac{1}{r_2}\right)+\frac{e^2}{r_{12}}.$$

The minimum of the first four terms is given by the spread of P_{r_1}, P_{r_2}, r_1, and r_2. The minimum of the last term is given by the largest value of r_{12} which must be less than r_1+r_2. The last limit requires that the three particles should lie on a straight line which is hardly realizable. However, the inequality

$$E > \frac{\hbar^2}{2m}\left(\frac{1}{\Delta r_1^2}+\frac{1}{\Delta r_2^2}\right)-ze^2\left(\frac{1}{\Delta r_1}+\frac{1}{\Delta r_2}\right)+\frac{e^2}{\Delta r_1+\Delta r_2}$$

is certainly true. It sets a lower limit to the actual minimum. Minimizing the above expression by varying Δr_1 and Δr_2, we get

$$E > -2(z-\tfrac{1}{4})^2 Rh.$$

In the following Table the limits thus calculated are given together with the experimental minima for the first few atoms.

	H^-	He	Li	Be	B	C
E (cal.)	−1·125	−6·125	−15·12	−28·12	−45·12	−66·12
E (exp.)	−1·05	−5·807	−14·56	−27·56	−44·06	−64·8

The discrepancy between the theoretical and the experimental values is to be expected from the crudeness of our estimate. The significant point is that all the calculated values lie lower than the empirical ones. The agreement can be much improved by considering the uncertainty in the electronic distributions. Either by averaging over all directions, remembering that r_{12} must be greater than either Δ_{r_1} or Δ_{r_2}, or by assuming correlated normal distributions for the electrons, we obtain, practically in all cases, results in reasonably good agreement within 1/2 to 1/5 per cent, with the experimental values.

Other complicated systems, such as Li-like and Be-like atoms, have been treated in the same way with reasonably good results, by assuring Pauli's principle and using the relation $\Delta P_r \Delta r \geqslant 2\hbar$ for the third and fourth particles (see footnote). It is, however, needless to go into details until the physical ideas presented here, for instance, the argument in the third example, can be formulated on a mathematical basis capable of extension to many-body problems.

I wish to thank Mr. W. T. Soh for some of the earlier calculations on the Berylium system and Dr. T. M. Hu for much of the later calculations.

December 18, 1949.

LXXVII. *Notices of New Books and Periodicals received.*

Radiative Transfer. By S. CHANDRASEKHAR. (Oxford: The Clarendon Press.) [Pp. 393.] Price 35*s*.

THIS book presents a systematic treatment of transfer in a plane parallel scattering medium, a subject which has been investigated principally by astrophysicists, but acquires new applications in the scattering of neutrons: indeed, other applications will assuredly be found in chemistry and applied physics for the quantitative results now available. The main line of attack consists in using certain principles of invariance (as that the diffuse scattering from a semi-infinite plane parallel atmosphere should be invariant to the addition of a layer of arbitrary optical thickness) to formulate integral equations: a method introduced by Ambarzumian in 1943 and greatly developed by Chandrasekhar himself. Astonishingly, the resulting non-linear inhomogeneous integral equations are then solved. Those who would recoil from the more difficult problems which are treated with success here, or who take no interest in radiative transfer, may nevertheless study with profit the application of Gaussian quadrature (and its generalization) to the approximate (and sometimes exact) solution of certain simpler integro-differential equations.

It is presumed that this will remain for long the standard text on its subject. But even should its techniques be superseded, the easy and illuminating style, and the continuous courtesy to the reader—in keeping him informed of where and how an argument is leading; in converting the plain " list of references " at the end of each chapter into concise and instructive " bibliographical notes "; in providing an index of definitions in addition to that of names and subjects—will allow it to stand as a model of the way in which a text in mathematical physics should be written.

F. C. F.

Absorption Spectrophotometry. By G. F. LOTHIAN, M.A., F. Inst.P. (A Hilger Publication.) Price 26*s*. net.

THIS book caters for the reader with a scientific background, commencing research in absorption spectrophotometry or requiring an informed survey of the practice and possibilities of the subject presented within reasonable compass. It succeeds admirably in these aims, and at the same time will be valued by experienced workers in the field as a balanced account of up-to-date practice.

The book is divided into three parts:—The first is introductory and assesses conditions affecting accuracy of measurement, and here, as throughout the book, the mathematics is an absolute minimum.

The second part, and perhaps the most valuable, presents a most interesting and varied collection of the types of problem arising in modern practice with appropriate methods of attack. Few readers will be familiar with the whole field surveyed, but specialists and non-specialists alike will profit by this comprehensiveness.

The third part discusses apparatus and techniques. The descriptions of apparatus are most effective, but those of actual techniques somewhat compressed.

The book is made up from good paper, is well illustrated with 70 figures and photographs, and also lists some 300 references, well indexed, and half of them within the 1944–48 period. This ample documentation mitigates those accounts which are necessarily brief in order to keep the book to its present length.

The reviewer finds a score of minor errors; a more careful layout of the position of the figures with reference to the text and a greater consistency in lettering would not be amiss. However, this is a good work and pleasant to read.

Cosmic Rays. By L. JÁNOSSY. 2nd Edition. [Pp. 454, Plates 10.] (Oxford: Clarendon Press, 1950.) Price 40*s*. net.

THE first edition of this work was published only two years ago and was reviewed in this Journal (**39**, 826, 1948) as authoritatively written by an investigator who has himself made numerous contributions to the subject. The need for a new edition at so early a date brings with it difficulties for the author, because important contributions to the subject are continuing to appear in rapid succession through the application of the photographic emulsion technique. It was probably inevitable at this stage that the attempt to bring the work up to date had to be made by what one may term tinkering and repair rather than by re-writing. But when a third edition is planned, changes in construction and in emphasis will undoubtedly be imperative. A.M.T.

History of Physics. By MAX VON LAUE, translated by R. E. Oesper. [Pp. 150.] (Academic Press, New York.) Price $2.30.

IN a foreword to his book, Professor von Laue says that the first draft was a war product written in 1943. "Precisely because of this circumstance the thought of the culture that is the common property of all nations, and which was then being so despicably mistreated, was ever present throughout the period of composition". A very readable book by an eminent physicist.

Manual of Organic Chemistry for Advanced Students, Vol. I. By G. DYSON. [Pp. 984.] (Longmans Green.) Price 63*s*.

THIS is the first part of a three volume treatise and deals with the compounds of carbon, hydrogen, oxygen and the halogens. Its scope may be indicated by saying that it is more comprehensive than a text-book, without being a collection of specialized monographs. Compiled by one who has made very important contributions in the field of notation in organic chemistry it may be looked upon as authoritative. Readers of this journal should find the volume useful as a work of reference.

BOOKS RECEIVED

The Association of Research Institutes of Tohuku University Japan have started to issue a series of Science Reports consisting of collected papers on work carried out by them in Physics Chemistry and Metallurgy. Volume I. Numbers 1 and 2 (May and August 1949) have been received.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LXXVIII. *Calibration of Proportional Counters by the Excitation of Flourescence Radiation with Radioactive Sources.*

By G. M. Insch,
Department of Natural Philosophy, Glasgow University*.

[Received July 3, 1950].

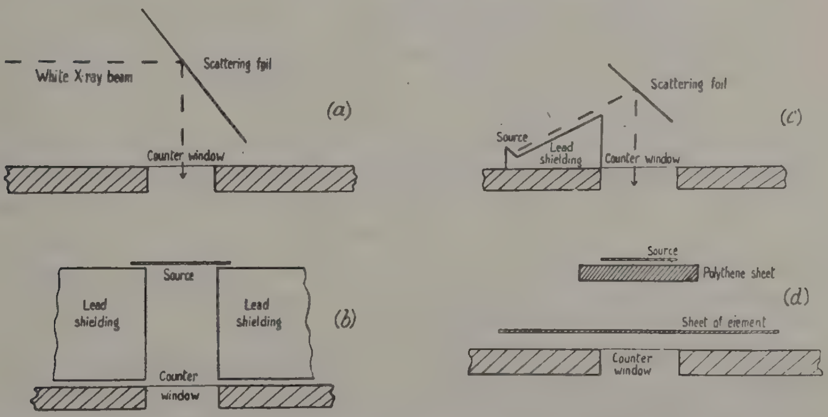
Summary.

The calibration of proportional counters by means of a very few suitably chosen radioactive sources which are used to excite the fluorescence radiation of various elements is described. The efficiency of excitation is high and the method is of wide application. The accurate identification of isomeric transitions using proportional tubes is demonstrated.

Introduction.

The Investigations carried out in this laboratory using proportional counters (*Phil. Mag.* 1949) have necessitated detailed examination of various calibration techniques. Information on this subject has already been published by Rothwell and West (1950). However, during the course of our work considerable progress has been made in overcoming the various problems involved, and it is the aim of this paper to correlate some of the results obtained. The chief method described here is original and it is described in reasonable detail since it is of wide application.

Fig. 1.


White X-ray beam
Scattering foil
Counter window
(a)
Scattering foil
Source
Lead shielding
Counter window
(c)
Lead shielding
Source
Lead shielding
Counter window
(b)
Source
Polythene sheet
(d)
Sheet of element
Counter window

In the early investigations the beam from an ordinary A.C.-operated X-ray machine, adjusted to operate at a small output intensity and which emitted white radiation was allowed to fall on a thin foil and the scattered radiation was detected in the counter at an angle to the beam direction (fig. 1(*a*)), (Curran, Angus and Cockroft 1949). This radiation entering the tube consisted mainly of the fluorescence X-rays of the scattering element. However, this method is not ideal. The X-ray machine is cumbersome, may give rise to spurious pulses (due to electrical interference), and calibration is accompanied by an increase in the " background " of the counter due to the detection of the white radiation. In addition the tube in ageing fails to self-rectify adequately and hence the pulses appear grouped on part of alternate half-cycles of the mains supply. For these reasons, new calibration techniques were explored and normally preferred. Pile-activated sources, which decayed by K-capture, or which showed nuclear isomersion or strong internal conversion of their γ-radiation, were found suitable as agents for the production of beams of homogeneous X-radiation or electrons. Sometimes the sources were used directly, but two methods of greatly entending their usefulness were established. The sources which have been most commonly applied in the course of several studies are shown in Table I.

---

* Communicated by the Author.

TABLE I.

Isotope	Half life	Transition	Approx. energy of X-ray	Associated radiation
In^{114}	50 days	isomeric	24·1 KeV.	1·98 MeV. $\beta+\gamma$'s
Sn^{113}	105 days	K-capture	24·1 KeV.	γ's and e^-
Ge^{71}	11 days	K-capture	9·2 KeV.	negligible
Fe^{55}	4 years	K-capture	5·9 KeV.	none if pure.

ELIMINATION OF THE EFFECTS OF HARD BETA AND GAMMA RADIATION.

The sources should have a reasonable half life for prolonged use. They should preferably be thin, particularly when associated with hard γ-radiation. Experience has shown that isotopes which have long period β-activity may be comfortably handled. The β-particles are absorbed by suitable thicknesses of polythene which has good transparency for X-radiation. For example, in the investigation of In^{114} (see below) the source was covered by a layer of polythene, 6 mm. thick. This completely absorbed the hard β-radiation (of energy 1·98 MeV.) and only reduced the X-radiation (energy 24·1 KeV.) by a relatively small fraction ($\sim$1/4). When the required soft calibrating X-radiation is associated with γ-radiation (of energy say $\sim$100 KeV.) some shielding is useful in reducing the undesirable γ-ray flux through the counter to a miniumm (fig. 1(*b*)). However, this method can only be applied with advantage to relatively soft γ-rays. In the case of still harder γ-radiation ($\sim$·25 MeV. upwards) the source is most conveniently examined when placed really close to the counter window. This simple geometry results in the best solid angle for soft radiation passing into the counter and makes use of the higher efficiency of the counter for such radiation. Usually it is possible to secure a ratio of at least 10 : 1 in efficiency of detection in favour of the soft X-radiation. To obtain optimum results when examining the source in this way, care must be taken to eliminate the " end-effect " of the tube, since a small fraction of the radiation detected penetrates to the region of variable gas gain.

PRODUCTION OF MODIFIED RADIATION.

Calibrating agents which gave homogeneous X-radiation covering the whole range from say 3 KeV. to 50 KeV. in small energy increments were found extremely useful. Since this coverage cannot be accomplished readily by use of radioactive sources alone new methods were devised. These involved the excitation of the desired radiation by homogeneous X-rays of higher energy originating from a radioactive source. This process is one of high efficiency and it was achieved in practice either by allowing the initial radiation to fall on a thin sheet of the appropriate element and detecting the excited radiation emitted from the front face (reflection (fig. 1(*c*)), or by placing a suitable thickness of the element between the source and the window (transmission). In the former case, the fluorescence radiation greatly predominates in intensity within the counter and its homogeneity is nearly independent of the thickness of the reflecting sheet. In the second case (transmission) a suitable thickness of the sheet can readily be calculated. For example, consider a source emitting homogeneous X-rays of wavelength λ_1 exciting a radiation of wavelength λ_2 in a sheet of the chosen element placed between the source and the counter window. Let this sheet have mass absorption coefficients μ_1 and μ_2 corresponding to λ_1 and λ_2. Let it be of thickness l and density p.

It can be shown that the number of fluorescence X-ray quanta entering the counter (fig. 1(*d*)) when I_0 quanta fall on the sheet is given by

$$N=K_1I_0[e^{-K_2}-e^{-K_1}]/(K_1-K_2),$$

where

$$K_1=\mu_1 pl \quad \text{and} \quad K_2=\mu_2 pl.$$

The number of unmodified X-rays transmitted through the sheet to the counter is

$$T=I_0e^{-K_1}$$

and the ratio of these is

$$R=N/T=K_1[e^{K_1-K_2}-1]/(K_1-K_2). \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Examination of equation (1) shows that the ratio R can be made as large as required (>100 say) by using sufficient thickness l, provided that $\mu_2<\mu_1$, a condition which is always satisfied in the usual experimental arrangements. (Furthermore this argument is strengthened by the fact that the modified radiation is usually the more efficiently detected). However, in practice large l values may necessitate strong sources and in some cases it may be difficult to secure adequate intensity. On the other hand, if $R\sim1$ is regarded as suitable (peaks due to modified and unmodified radiations of equal intensity) the method can be used with very weak sources. (see fig. 2).

Regarding the reflection method (fig. 1(*c*)) it is sufficient to note that it can be readily applied in almost every case, but it may be inferior to the transmission method when very weak sources are employed.

These methods were applied to excite the fluorescence radiations of a number of elements (see Table II.). In this Table E_0 is the approximate energy of the initial X-radiation falling on the element and E_m is the approximate energy of the modified radiation.

With the simple geometry demonstrated in fig. 1(*c*) it was possible to detect a) modified radiation ~5 per cent of the total radiation emanating from the source. Hence a source strength of $<1\mu$ curie total activity proved sufficient. Indium and Geranium sources were found adequate for excitation of the flourescence radiations from less than 5 KeV. upwards, which considerably increased the number of homogeneous X-ray calibrating agents available within this energy range.

Fig. 2.

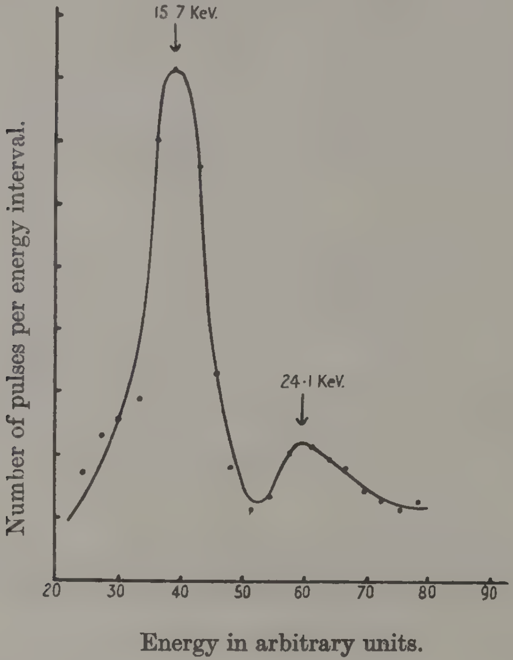


Table II.

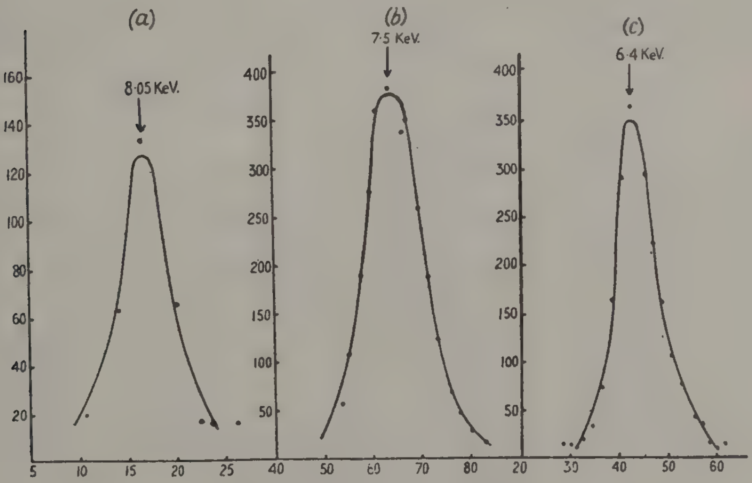
Source	Excited element	E_0 in KeV.	E_m	Type of geometry	Histogram
In^{114}	Zr	24·1	15·7	Tramsnission	fig. 2
In^{114}	Cu	24·1	8·05	Reflection	fig. 3(*a*)
Ge^{71}	Ni	9·2	7·5	Reflection	fig. 3(*b*)
Ge^{71}	Fe	9·2	6·4	Reflection	fig. 3(*c*)

New calibrating agents are at present being examined for use with proportional counters in magnetic fields. Curran, Cockroft and Insch (1950) have recently used proportional tubes to measure particle energies >1 MeV. These examinations demand calibration radiations of several hundred KeV. energy. Sources which decay by γ-ray transitions (not associated with β-activity) and which show strong internal conversion, are being studied (*e.g.* Sn^{113}; γ-ray energy 392 KeV. and $\sim$70 per cent internal conversion) Such sources are introduced into the counter.

IDENTIFICATION OF ISOMERIC TRANSITIONS.

The primary phenomena externally observable in an isomeric transition are the emission of γ-radiation and usually conversion electrons. Accompanying the latter is a K X-radiation which is detected with high efficiency and accuracy by a proportional counter. Since this X-radiation is frequently difficult to identify with certainty, the method described here is an important aid. The isomeric transition in In^{114} was studied and the Kα fluorescence X-ray histogram is shown in fig. 4, curve A. At the same time, the Kα fluorescence radiations of cadmium (energy 23·1 KeV.) Indium (24·1 KeV.) and tin (25·2 KeV.) were excited and compared with the radiation arising in the nuclear process. The three histograms are shown in fig. 4, curves B, C and D. This demonstrates at once that the known γ-ray energy 192 KeV., which is strongly converted, is emitted prior to any transition in which Z changes. The importance of this method for the accurate identification of isomeric processes is obvious.

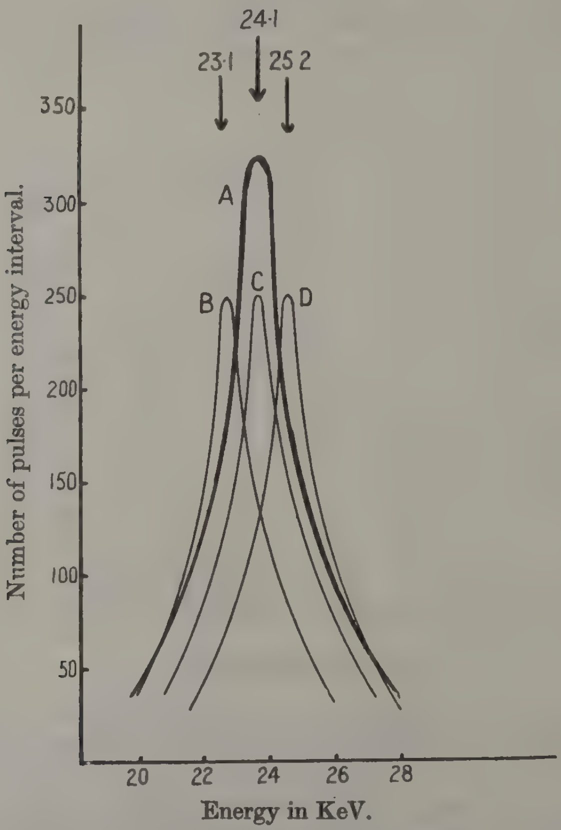
Fig. 3.



Energy in arbitrary units.

In the course of this work argon was used as the absorbing gas. For the energies investigated it is superior to the gases of greater atomic number, such as Xenon and Krypton, which have a relatively small Auger effect, causing homogeneous X-radiation to yield two peaks of comparable intensity, which can lead to uncertainties. One peak corresponds to the case when the argon X-ray escapes from its own atom and is not detected. The Auger effect in argon is high and an independent value for this effect was obtained during the work described above. This was accomplished by studying an X-ray histogram obtained when the fluorescence K X-rays of iron were excited by a germanium source. Two peaks were evident, one due to the fluorescence Kα X-radiation of iron of energy 6·4 KeV., and one to the photoelectrons (energy 3·4 KeV.) from the K level of argon when the X-ray escaped and did not form an Auger electron.

Fig. 4.



The effect on the smaller energy peak, due to 3 KeV. X-radiation caught in the gas was calculated and the " background " spectrum of the counter was subtracted from the histogram. Using this the Auger effect in argon was found to be 94 per cent in reasonable agreement with previous work (Compton and Allison 1936).

Throughout this work a counter $5\frac{1}{2}''$ internal diameter was used with a filling of 60 cm. Hg. of argon and 15 cm. Hg. of methane.

References.

Rothwell, P., and West, D., 1950, *Proc. Phys. Soc.*, **63,** 541.
Curran, S. C., Angus, J., and Cockroft, A. L., 1949, *Phil. Mag.*, **40,** 36.
Curran, S. C., Cockroft, A. L., and Insch, G. M., in the press.
Compton, A. H., and Allison, S. K., 1936, *X-rays* (Macmillan).

LXXIX. *Transit-time Phenomena in Electron Streams.*—II.

By D. K. C. MacDonald,
Clarendon Laboratory, Parks Road, Oxford *.

[Received July 13, 1950.]

Abstract.

A density modulated electron stream will suffer amplitude reduction and phase variation with transit time due to the statistical velocity distribution of the electrons. The problem is considered in relation to that of noise generation in a beam due to velocity fluctuation; the analysis developed is then used in examining the variation of space-charge reduction factor (Γ^2) under particular laws of potential distribution which present themselves in physical problems.

§1. Introduction.

In a previous paper in this Journal † (MacDonald 1949 a) we analysed the problem of an electron beam initially accelerated to a relatively high potential and initially ordered in density ($\Gamma^2=0$), which becomes progressively disordered ($\Gamma^2 \to 1$) due to the random distribution of velocities present in the electron beam. We wish now to consider the related problem of an electron stream with thermal velocities which has a given density-modulated sinusoidal "signal" initially imposed; the amplitude and phase of this signal will then suffer alteration with transit-time due to the thermal energy spread. We shall examine the significance of this relationship and then proceed to analyse the variation of such a signal and the related Γ^2 under certain potential laws of physical interest. The theoretical predictions will be compared with such experimental evidence as is available and the significance of the results discussed.

§2. Reduction of Signal Amplitude.

Since (I) was published, both van der Ziel (1949) and MacDonald (1949 b) have independently considered the variation during drift time τ of a sinusoidal density-variation in an electron stream of thermal energy distribution

$$\left.\begin{aligned} p(\mathrm{E})d\mathrm{E} &= 1/k\theta \,.\, \exp(-\mathrm{E}/k\theta)d\mathrm{E}, \qquad (\mathrm{E}>0) \\ &=0 \qquad\qquad\qquad\qquad\quad , \qquad (\mathrm{E}<0) \end{aligned}\right\} \qquad (1)$$

(k : Boltzmann's constant. θ : Absolute temperature)

which has been initially accelerated very rapidly to energy Φ.

If the beam-density before "drifting" ($\tau=0$) is given by

$$n_{(0)}=\mathrm{N}_0+\mathrm{A}_0 \sin \omega t,$$

---

* Communicated by the Author.

† Referred to as (I) here,

then one shows readily by considering the loss and gain of electrons in a short interval of time that

$$n_{(\tau)}=\frac{2\Phi}{k\theta\tau}\int_0^\infty \exp\left(-\frac{2\Phi x}{k\theta\tau}\right) . (\mathrm{N}_0+\mathrm{A}_0 \sin \omega t \cos \omega x+\mathrm{A}_0 \sin \omega x \cos \omega t)\, dx$$

$$=\mathrm{N}_0+\mathrm{A}_0 \sin \omega t\left\{\frac{1}{1+a^2}\right\}+\mathrm{A}_0 \cos \omega t\left\{\frac{a}{1+a^2}\right\}, \quad . \quad . \quad . \quad . \quad (2)$$

where $$a=\omega t\left(\frac{k\theta}{2\Phi}\right).$$

The unvarying (" d.c. ") component, N_0, remains unchanged, as we should expect, but the oscillatory (" a.c. ") component suffers reduction in magnitude and a change in phase. It is perhaps interesting to observe that for the large values of τ the phase-shift approaches $\pi/2$ asymptotically.

§ 3. Relationship with Noise-Reduction Factor, Γ^2.

Equation (2) is in itself of interest ; by comparison with (I), however, where we found $\Gamma^2=\frac{a^2}{1+a^2}$, we may show immediately

$$\mathrm{A}_\tau=\mathrm{A}_0 \sqrt{(1-\Gamma^2)} \sin (\omega t+\alpha), \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\left[\alpha=\tan^{-1}\sqrt{\left(\frac{\Gamma^2}{1-\Gamma^2}\right)}\right],$$

where A_τ is the signal after drift time, τ. The variation of the signal components and of Γ^2 are plotted in fig. 1, particularly for comparison with the later work.

Equation (3) may be interpreted as follows. Consider a wholly disordered stream (full " shot effect ", $\Gamma^2=1$); then the noise current $\overline{\delta \mathrm{I}^2}$ may be considered as distributed in a frequency spectrum according to the well-known expression

$$\overline{\delta \mathrm{I}_f^2}=2e\mathrm{I}\, df. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

After drifting for a time τ, according to (3), the magnitude will have diminished to

$$\overline{\delta \mathrm{I}_{f,\tau}^2}=2e\mathrm{I}(1-\Gamma^2)\, df. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

On the other hand, the d.c. component, I, which we have effectively considered as " smooth " by distributing the fluctuation throughout the frequency spectrum in (4), will, by (I), have become partially disordered according to the law

$$\overline{\delta \mathrm{I}^2}=2e\mathrm{I}\Gamma^2\, df. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4b)$$

The net sum of (4a) and (4b) is thus simply $2e\mathrm{I}df$. The irreversible maintenance of full shot effect (equation (4)) in the beam may thus be regarded as the " detailed balancing " of these two mechanisms always proceeding at the same rate at any instant. When we consider the problem, as in (I), of the beam with zero shot effect, then only one process can occur and we thus witness the so-called irreversible approach towards full shot effect *. We remark also that the *signal-noise (power) ratio* will evidently deteriorate following a law $\sim \frac{1-\Gamma^2}{\Gamma^2}$.

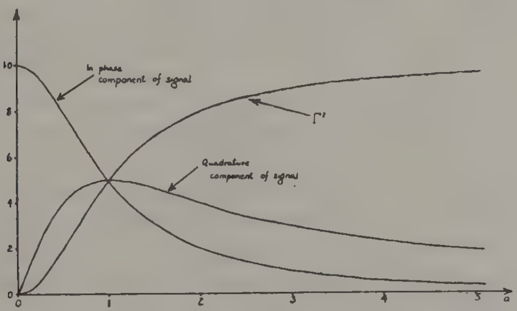
With the equivalence demonstrated above the calculation of Γ^2 in certain problems is simplified.

§ 4. LEMMA—CASE OF INITIAL $\Gamma^2 \neq 0$.

Before continuing to a detailed analysis of the problems in view we state a simple lemma. Thus far we have assumed that $\Gamma^2=0$ initially; we ask now how the noise will grow if initially

$$\overline{\delta I_f^2}=2eI\Gamma_0^2\,df \qquad (0<\Gamma_0^2\leqslant 1).$$

Fig. 1.



Variation of signal components and Γ^2 for accelerated electron-stream.

If $\Gamma^2_{\tau,0}$ be the noise factor due to transit calculated for an initially smooth beam ($\Gamma_0^2=0$), then the overall factor (say Γ_τ^2) is given by

$$\Gamma_\tau^2=\Gamma_0^2+(1-\Gamma_0^2)\Gamma^2_{\tau,0}. \qquad (5)$$

A simple demonstration may be given by splitting the current, I say, into an initially "noisy" component of magnitude $I\Gamma_0^2$ and an initially "smooth" one of magnitude $I(1-\Gamma_0^2)$ (*cf.* (I), p. 565).

§ 5. Γ^2 IN VARIOUS POTENTIAL DISTRIBUTIONS †.

In (I) we assumed that the electron beam was injected into the drift region at relatively high energy, Φ, with $\Gamma^2=0$. In practice the electron stream is derived from a structure ("electron gun") similar in principle to a simple thermionic diode, of anode voltage Φ/e (measured from the potential minimum, strictly speaking), having a hole in the anode for projection of the electron beam. The smoothing in density of the emitted current fluctuations occurs through the medium of the variable potential minimum which is present between cathode and anode when the current is limited by space-charge. At medium frequencies the overall transit time in the electron gun is so short that a quasi-steady state may be assumed in the "classical" analysis (*cf.* North (1940), Schottky and Spenke (1937)); consequently one need not consider what region of the "valve" is dynamically responsible for the smoothing process. At very high frequencies, however, the very existence of a high degree of smoothing in diodes and long beams (Kompfner *et al.* 1946; Kompfner 1947) suggests that the action must occur effectively within a very small fraction of the total diode transit time. Further, the general assumption that the "compensating" stream of electrons released is of almost zero energy (relative to the space-charge minimum) suggests that essentially only a small region about the potential minimum is significant for the initial smoothing.

---

* v. d. Ziel's communication (*loc. cit.*) is on somewhat similar lines to the foregoing. He suggests, however, that Γ^2 is derivable *à priori* from an equation equivalent to (2) of this paper. We have reason to prefer to regard the two analyses as of *individual* significance, demonstrating also, however, *under these assumptions* (*e. g.* no space-charge interaction) the equivalence shown above.

† It should perhaps be remarked that we are here throughout interested in the dynamics of an electron stream—that is, in the convective part of the current. The experimental data used below for comparison refer indeed to such currents.

As a first analytical model, then, on the basis of these considerations, we shall assume that the space-charge smoothing is effectively complete immediately beyond the potential minimum. We would emphasize that no theory as yet exists for the detailed space-charge interaction of electrons in a valve at very high frequencies; we hope in the future to be able to make some progress in this direction. If a satisfactory detailed theory is forthcoming we should then be able to specify with greater certainty the "initial conditions" of our present problem.

We turn thus to consider the current fluctuations in the electron gun itself. We measure potential, V, transit time, τ, and distance, x, from the potential minimum. In line with the foregoing discussion, we assume first a smooth current "injected" at $x=0$; this corresponds to the assumption that the smoothing mechanism can be idealized, in the first place, to occur at $x=0$.

(*a*): *Linear potential variation* *.

Let $V=Cx$; we first require the law of transit-time, τ_E, from $V=0$ at $x=0$ to potential V_0 at $x=x_0$, say, with variation of initial (thermal) energy, E.

Thus

$$\tau_E=\sqrt{\left(\frac{m}{2e}\right)}\int_0^{x_0}\frac{dx}{\sqrt{(V+E/e)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$=\tau_0(1-(E/eV_0)^{1/2}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where τ_0 is transit time for $E=0$. Equation (6) then replaces the equation for τ appropriate in (I) and in § 2 of this paper, namely

$$\tau_E=\tau_0(1-E/2\Phi).$$

---

* It is of course understood that a linear law could not exist fundamentally in the immediate neighbourhood of the minimum; it is chosen simply as an obvious overall first approximation.

To calculate Γ^2 we rely on the equivalence demonstrated in §3 above. We find in this case that $S_\tau \equiv A_\tau/A_0$ is given by

$$S_\tau = \sin \omega t \int_0^\infty \epsilon^{-x^2} \cos\left(\frac{\omega x}{\beta}\right) 2x\,dx + \cos \omega t \int_0^\infty \epsilon^{-x^2} \sin\left(\frac{\omega x}{\beta}\right) 2x\,dx, \qquad (8)$$

where $$\beta = \frac{1}{\tau}(eV_0/k\theta)^{1/2},$$

assuming that the electrons leave the minimum, as is well-known, with a thermal energy distribution given by equation (1) above.

Thus

$$S_\tau = \sin \omega t \left\{1 - a\int_0^\infty \epsilon^{-x^2} \sin ax\,dx\right\} + \cos \omega t \left\{a \int_0^\infty \epsilon^{-x^2} \cos ax\,dx\right\} \qquad (9)$$

where $a = \omega\tau(k\theta/eV_0)^{1/2}$ now.

$$= \sin \omega t \left\{1 - a\epsilon^{-a^2/4} \int_0^{a/2} \epsilon^{t^2}\,dt\right\} + \cos \omega t \left\{\frac{a\sqrt{\pi}}{2}\epsilon^{-a^2/4}\right\}. \quad . \quad . \quad . \quad . \qquad (10)$$

Hence $$\Gamma^2 = 1 - \left\{\left(1 - a\epsilon^{-a^2/4}\int_0^{a/2} \epsilon^{t^2}\,dt\right)^2 + \frac{a^2\pi}{4}\epsilon^{-a^2/2}\right\}. \quad . \quad . \quad . \qquad (11\,a)$$

The function $\int_0^x \epsilon^{t^2}\,dt$ is tabulated (*e. g.* Jahnke and Emde (1933), p. 32), or Lash Miller and Gordon (1931) tabulate $\epsilon^{-x^2}\int_0^x \epsilon^{t^2}\,dt$. One may also usefully use expansion in powers of a, showing in particular that for small a

$$\left.\begin{aligned} \Gamma^2 &\doteqdot a^2\left(1 - \frac{\pi}{4}\right) + \frac{a^4}{24}(10 - 3\pi) + \dots \\ &= 0{\cdot}215a^2 + 0{\cdot}024a^4 + \dots \end{aligned}\right\} \quad . \quad . \quad . \quad . \qquad (11\,b)$$

The relevant functions in (10) and (11*a*) are plotted in fig. 2, with a as abscissa. We observe now that a phase change of π in the signal occurs as $\tau \to \infty$.

In comparison with experiment, we may take the following values for a typical electron gun used at very high frequencies :

$f\tau$ (number of transit cycles) ≈ 3

$V_0 = 4 \times 10^2$ volts

$\theta \simeq 10^3$ °K. $\therefore \frac{k\theta}{e} \simeq 8{\cdot}6 \times 10^{-2}$ volts.

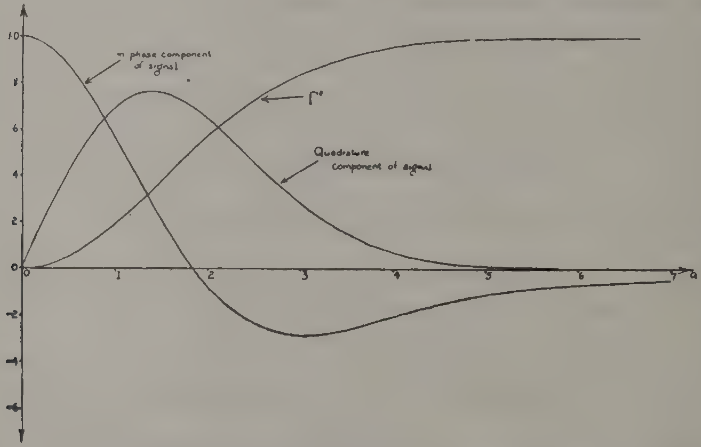
Hence, in this case, $a = 6\pi\sqrt{(2{\cdot}15)} \times 10^{-2} \simeq 0{\cdot}28$, and therefore

$$\Gamma^2 = 0{\cdot}016_5. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (12)$$

Since experimental work based on the beam derived from such a gun (*cf*. Robinson and Kompfner 1950) indicates $\Gamma^2 \sim 0{\cdot}02$ this theoretical result might be considered most satisfactory, suggesting that the conditions postulated ($\Gamma^2=0$ at $x=0$, etc.) could be regarded as providing an acceptable model for the fluctuational behaviour at high frequencies.

Before accepting this indication, however, it appears essential to consider the effect of closer approximation to the true potential distribution in the valve "gun".

Fig. 2.



Variation of signal components and Γ^2 when $\mathrm{V}=\mathrm{C}x$.

(*b*) *The Langmuir–Child Potential Law.*

Child (1911) and Langmuir (1913) showed that the current, I, in an idealized parallel-plane diode, limited solely by space-charge, initial cathode emission velocities being neglected, is given by the now classical expression

$$\mathrm{I}=\frac{1}{9\pi}\sqrt{\left(\frac{2e}{m}\right)}\frac{\mathrm{V}_a^{3/2}}{d^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where d: anode-cathode distance.

V_a: anode-cathode potential difference.

We adopt this diode as a second model for the potential-minimum/anode region in the electron gun structure, where the minimum is now considered as a virtual cathode at $x=0$. The law of potential variation with x corresponding to (13) is then

$$\mathrm{V}=\mathrm{C}x^{4/3}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where C is a constant for the structure under a particular current.

The transit-time law with initial energy E (*cf*. equation (6) above) is then

$$\tau_{\mathrm{E}}=\sqrt{\left(\frac{m}{2e}\right)}\int_0^{x_0}\frac{dx}{(\mathrm{C}x^{4/3}+v)^{1/2}}, \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where $v=\mathrm{E}/e$, the equivalent initial voltage.

Simple substitution transforms (15) to

$$\tau_{\mathrm{E}}=\frac{2}{3}\left(\frac{v}{c}\right)^{1/4}\sqrt{\left(\frac{m}{2e\mathrm{C}}\right)}\int_{0}^{\sinh^{-1}(x_0^{2/3}\sqrt{(c/v)})}\sinh^{1/2}\phi\,d\phi,$$

which after expansion, integration and some reduction yields

$$\tau_{\mathrm{E}}=\tau_0\{1-0{\cdot}85(v/\mathrm{V}_0)^{1/4}+0{\cdot}167(v/\mathrm{V}_0)-0{\cdot}038(v/\mathrm{V}_0)^2+\ldots\}. \quad . \quad (16)$$

An alternative, more perspicuous, approach is to split the range of integration in (15) at $x=\xi$, where ξ is the value of x for $\mathrm{V}=v$. The two integrals may then be expanded and we obtain

$$\left.\begin{aligned}\tau_{\mathrm{E}}(1)&\approx\tau_0(0{\cdot}28(v/\mathrm{V}_0)^{1/4}),\\ \tau_{\mathrm{E}}(2)&\approx\tau_0(1-1{\cdot}13(v/\mathrm{V}_0)^{1/4}+0{\cdot}167(v/\mathrm{V}_0)\ldots).\end{aligned}\right\}\quad . \quad . \quad . \quad (17)$$

These expressions themselves seem not without interest in view particularly of a recent discussion by Pierce (1948) of transit-time variation with initial energy in a valve with a potential minimum due to space charge. He measures time, t, as we do, from the place of zero field (minimum of potential) and then uses the equation

$$\frac{\partial \mathrm{V}}{\partial x}=4\pi \mathrm{I}t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

(his equation (43))

to determine the transit time.

Where I is the diode current, and t the transit time of an electron to x. This equation, however, as an application of Poisson's law—remembering $\partial\mathrm{V}/\partial x\equiv 0$ at the minimum—is only valid where all electrons (producing I) have the *same* velocity and t is the transit time of one of them and it is thus quite inappropriate in the present situation. If, indeed, I be supposed produced by a stream of electrons leaving the minimum with zero velocity and t be the transit time of a "perturbed" electron with positive initial velocity then it is obvious that the error in (18) is gross indicating in fact a variation of $\partial V/\partial x$ in the wrong sense.

Equations (16) and (17) are also interesting in view of the common assumption (*e. g.* Rack (1938, p. 613)) that *thermal* energies produce negligible transit time variation in a valve except under extreme conditions. If, however, we set

$$\mathrm{V}_0\sim 10^2 \text{ volts}; \qquad v=k\theta/e\sim 10^{-1} \text{ volts},$$

then from (16): $\tau_{\mathrm{E}}\doteqdot\tau_0(1-0{\cdot}15)$, showing that in only 3 cycles transit time the difference would already amount to $\sim\frac{1}{2}$ cycle.

Returning to our analysis we proceed as in (*a*), using now

$$\tau_{\mathrm{E}}\approx\tau_0(1-0{\cdot}85(v/\mathrm{V}_0)^{1/4}), \quad . \quad . \quad . \quad . \quad . \quad (16a)$$

in place of (6),

remembering also $(v/\mathrm{V}_0)\ll 1$. In this case we find that the relative signal is now

$$\mathrm{S}_\tau=\sin\omega t\left(1-a\int_0^\infty \epsilon^{-x^4}\sin ax\,dx\right)+\cos\omega t\left(a\int_0^\infty \epsilon^{-x^4}\cos ax\,dx\right) \quad . \quad (19)$$

and

$$\Gamma^2=1-\left\{\left(1-a\int_0^\infty \epsilon^{-x^4}\sin ax\,dx\right)^2+\left(a\int_0^\infty \epsilon^{-x^4}\cos ax\,dx\right)^2\right\} \qquad (20)$$

where now $a=\omega\tau\, 0{\cdot}85(k\theta/eV_0)^{1/4}$.

The integrals in (19) may again be expanded and we have

$$S_\tau=\sin\omega t\left(1-\frac{a^2\sqrt{\pi}}{4}+\frac{a^4}{4{\cdot}3\,!}\,\ldots\right)+\cos\omega t\left(\Gamma(5/4)\,.\,a-\frac{\Gamma(7/4)}{3{\cdot}2\,!}a^3\ldots\right), \qquad (21a)$$

whence for small a

$$\Gamma^2=a^2\left(\frac{\sqrt{\pi}}{2}-(\Gamma(5/4))^2\right)-a^4\left(\frac{\pi}{16}+\frac{1}{12}-\frac{\Gamma(5/4)\Gamma(7/4)}{3}\right)\ldots$$

$$\approx 0{\cdot}066a^2-0{\cdot}002a^4 \qquad (21b)$$

S_τ and Γ^2 could be calculated in fact as far as $a=2$ using such expansions but the work quickly becomes lengthy. For very large a, asymptotic expansions may be sought, but are not of great value here. It was, however, found on plotting accurately ϵ^{-x^4} that a very close approximation to this function can be found through the formula

$$\left.\begin{aligned} f(x)&=1, && 0\leqslant x\leqslant 0{\cdot}3\\ &=1-1{\cdot}36(x-0{\cdot}3)^2, && 0{\cdot}3<x<0{\cdot}91\\ &=1{\cdot}36(1{\cdot}52-x)^2, && 0{\cdot}91<x<1{\cdot}52\\ &=0. && 1{\cdot}52\leqslant x \end{aligned}\right\} \qquad (22)$$

Thus an explicit integration of (19) and (20) could be obtained:

$$S_\tau=\sin\omega t\left\{\cos 0{\cdot}91a\left(-0{\cdot}011+\frac{5{\cdot}44}{a^2}(1-\cos 0{\cdot}61a)\right)\right\}$$

$$+\cos\omega t\left\{\sin 0{\cdot}91a\left(-0{\cdot}011+\frac{5{\cdot}44}{a^2}(1-\cos 0{\cdot}61a)\right)\right\} \qquad (23a)$$

$$\Gamma^2=1-\left\{-0{\cdot}011+\frac{5{\cdot}44}{a^2}(1-\cos 0{\cdot}61a)\right\}^2$$

$$\doteqdot 1+\frac{0{\cdot}120}{a^2}(1-\cos 0{\cdot}61a)-\frac{29{\cdot}57}{a^4}(1-\cos 0{\cdot}61a)^2. \qquad (23b)$$

The values given by (23 a, b) agree excellently with those derived from (21 a, b) in the common region calculated ($a\leqslant 2$) and the relevant functions are plotted in fig. 3.

Turning to the numerical example quoted above, we now have $a=2{\cdot}175$, whence from fig. 3, $\Gamma^2=0{\cdot}26$. This figure is now greatly in excesss of the experimental value and it is further evident from the lemma of §4 that if $\Gamma_0^2\neq 0$ the situation is merely aggravated.

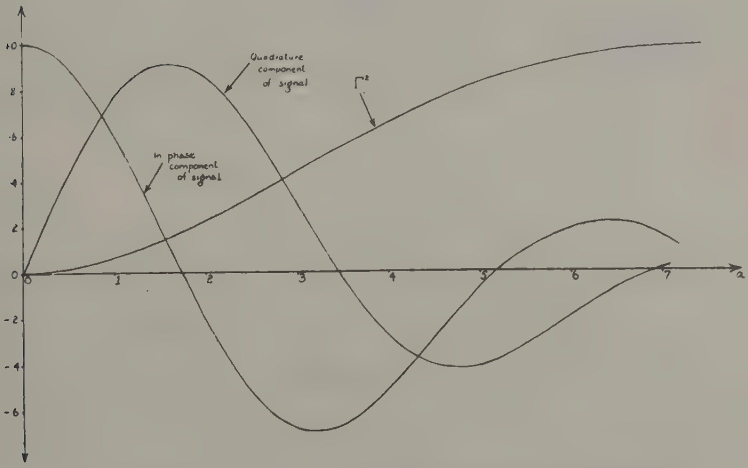
(c) *Discussion.*

It appears clear that this disagreement arises at least in part from our assumption that the smoothed current "emerges" from precisely $x=0$; $V=0$. In fact it is evident that the dynamical smoothing process cannot be *entirely* localized, while a glance at equation (17) shows clearly the important contribution to the above value for Γ^2 played by the potential in the immediate neighbourhood of the minimum. This feature is evinced even more strongly when we recollect that with the true velocity spectrum the potential *initially* rises with a square law (*cf. e. g.* Langmuir and Compton 1931), since in that case if we again postulate a smooth current from $x=0$, the initial rise of Γ^2 may be shown to be exceedingly rapid following a law

$$\Gamma^2 \sim \omega^2\tau^2 \log\frac{1}{\gamma\omega^2\tau^2}, \quad \dots\dots\dots \quad (24)$$

(γ, $\equiv e^c$ (where C is Euler's constant), $=1\cdot781\ \dots$).

Fig. 3.



Variation of signal components and F^2 when $V=Cx^{4/3}$.

On the other hand in such a vital region it then becomes imperative to consider other factors in the physics of the fluctuation problem such as the Coulomb repulsion of the electrons; at this point we recollect the lack of a detailed theory of space-charge interaction mentioned previously.

Returning, however, to the Langmuir–Child law the *significance* of this region may be readily assessed from fig. 4. The variation of Γ^2 for the numerical values previously laid down has been computed when it is assumed that the smooth beam "emerges" from a potential value *greater* than zero up to "thermal" potential, $k\theta/e$. In particular we find that in the latter case, Γ^2 has fallen as low as $\sim 0\cdot006$, well below the experimental value, restoring some reasonable confidence in the postulated model. It is probably significant also to notice that the transition from the quadratic potential law to a 4/3 power law occurs in the region $(1\rightarrow 10)\times k\theta/e$.

It is clear that in the linear approximation of (*a*) small weight attached to the very initial region and consequently the condition $\Gamma^2 \approx 0$ at $x=0$ was adequate.

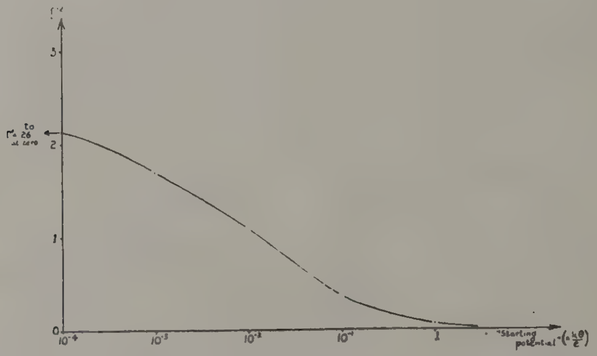
§ 6. CONCLUSION.

The analysis presented above cannot be regarded as necessarily final in nature, but it is offered in conjunction with the models discussed as a contribution to the study of the behaviour of electron streams.

§ 7. ACKNOWLEDGMENTS.

It is a pleasure to thank Mr. R. Kompfner for many stimulating discussions, Dr. Kathleen Sarginson for her help, and Professor M. H. L. Pryce for his interest in the work.

Fig. 4.



Variation of Γ^2 with " starting potential ", when $V=Cx^{4/3}$.

REFERENCES.

CHILD, C. D., 1911, *Phys. Rev.*, **32**, 492.
JAHNKE, E., and EMDE, F., 1933, *Funktiontafeln*, Teubner, Leipzig.
KOMPFNER, R., 1947, *Jour. Brit. Inst. Rad. Eng.*, **7**, 117.
KOMPFNER, R., *et al.*, 1946, *Jour. Inst. Elec. Eng.*, **93** (IIIA), 225.
LANGMUIR, I., 1913, *Phys. Rev.*, **2**, 450.
LANGMUIR, I., and COMPTON, K., 1931, *Rev. Mod. Phys.*, **3**, 244.
LASH MILLER, W., and GORDON, A. R., 1931, *Jour. Phys. Chem.*, **35**, 2878.
MACDONALD, D. K. C., 1949 a, *Phil. Mag.*, (7), **40**, 561 ; 1949 b, Unpublished SERL Memo : " Reduction of signal amplitude due to thermal fluctuation in long electron beams " (October 1949).
NORTH, D. O., 1940, *R.C.A. Review*, **4**, 269, 441.
PIERCE, R., 1948, *Bell Syst. Tech. Jour.*, **27**, 159.
RACK, A. J., 1938, *Bell Syst. Tech. Jour.*, **17**, 613.
ROBINSON, N., and KOMPFNER, R., 1950, In course of publication.
SCHOTTKY, W., and SPENKE, S., 1937, *Wiss. Veröff. Siemenswerken*, **16**, I.
V. D. ZIEL, A., 1949, *Proc. Inst. Rad. Eng.*, **37**, 1447 (Letter to Editor).

LXXX. *Proportional Counters in a Magnetic Field: an Investigation of the Isomerism of Br^{80} and a Measurement of the Fluorescence Yield of Krypton**.

By D. West and Miss P. Rothwell,
Atomic Energy Research Establishment, Harwell †.

[Received May 17, 1950].

Summary.

The proportional counter technique for measuring the energy of soft electrons has been extended to include higher energy radiations by placing the counter in a strong magnetic field. Suitable X-ray calibration sources consisting of long lived K-capture bodies which give lines at energies up to 50 keV. have been developed.

In an application of the technique to a study of the isomerism of Br^{80}, the conversion coefficient of the 37 keV. gamma ray was found to be 1·2, which is close to the calculated value for an electric dipole transition.

In a second application, the fluorescence yield of the K shell of Krypton was measured. The value obtained was 0·67.

§ 1. Introduction

The use of proportional counters for studying the energies of soft radiations has been described recently in a series of papers by Kirkwood, Pontecorvo and Hanna, and by Curran, Augus and Cockroft. For accurate measurements the great majority of ionizing particles must spend all their energy in the active volume of the counter; if a particle escapes from the counter before coming to rest, the size of pulse it produces will not correspond to its energy and a distorted energy spectrum will be obtained. The limit to the energy which can be measured with a proportional counter is therefore set by the dimensions of the counter and the pressure of gas inside it. In this paper it will be shown that this limit may be extended to higher energies for any particular counter by placing it in a strong magnetic field with its axis parallel to the field. The escape of particles through the side wall of the counter is reduced, since the track is coiled about an axis parallel to the field direction. By using a strong magnetic field, the diameter of a coiled track (for an electron moving perpendicular to the field) can be made much smaller than the range, as can be seen from Table I. in which electron scattering is neglected.

---

* Some of the results given here have already been reported briefly (Rothwell and West 1950.)

† Communicated by the Authors.

Of course, particles having a small velocity component perpendicular to the axis will escape through the ends of the counter even when the magnetic field is strong enough to prevent escape from the side walls. The escape from the ends, however, is always much less important than escape through the side walls. The effect of Coulomb scattering along the electron track complicates the situation. As far as escape through the side walls is concerned, the scattering will tend to make the lateral spread of any track somewhat greater than the radius of curvature given in the table and so will reduce the efficacy of the field slightly. Scattering will, however, assist in reducing escape of particles along the axis since it prevents any particles from moving nearly parallel to the field for long and as a result being uninfluenced by the field. Considerable reduction in the number of electrons escaping is therefore to be expected when a magnetic field is applied.

TABLE I.

Electron energy	Range in air in N.T.P.	Radius of curvature in magnetic field of 7500 gauss. (for an electron moving perpendicular to the field)
50 keV.	4 cm.	0.1 cm.
100 keV.	12 cm.	0·15 cm.
500 keV.	140 cm.	0·39 cm.

The possiblity that the multiplication properties of a counter might be affected by a strong magnetic field was considered. As discussed below it was found experimentally that the multiplication properties were only slightly affected by the magnetic field, in a way which did not invalidate the technique.

The technique has been tested in conditions where, in absence of the field, the electron escape would prevent energy measurements, Electrons of energies up to 50 keV. have so far been studied with this technique in investigations of the nuclear isomerism of Br^{80} and of the Auger effect in krypton.

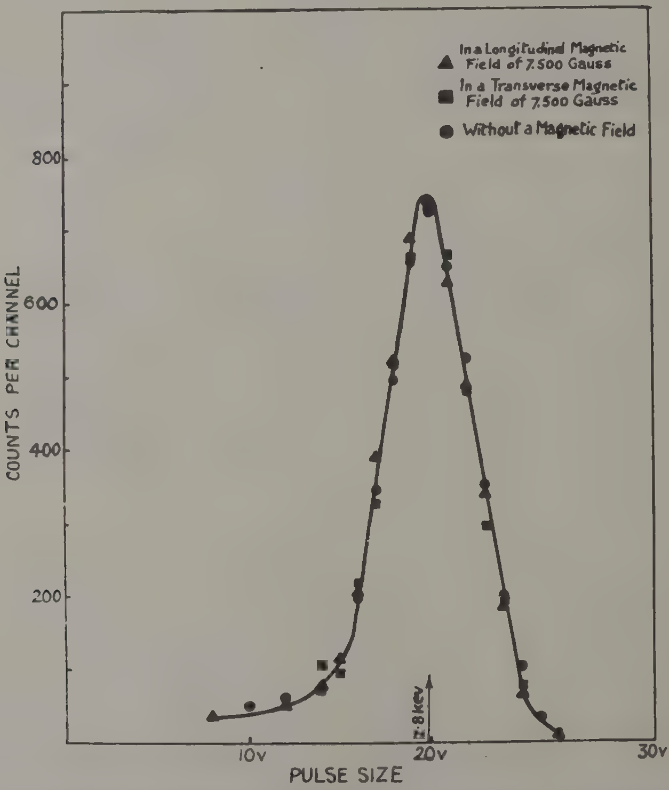
§ 2. DESCRIPTION OF APPARATUS.

The counters used in this work were made of soft glass with an aqua-dag cathode painted on the outside (Maze 1946). The wall thickness was 1 mm., the diameter 5 cm., and the active length was 34 cm. The central wire was tungsten 0·005 in., in diameter. This type of counter was used in preference to the conventional metal walled counter to reduce the chance of active gas being adsorbed on the walls. Each counter was fitted with a glass tap and a ground glass joint by means of which it could readily be connected to a filling apparatus for evacuation and refilling. Pulses were taken off the wire and fed through a head amplifier and a linear amplifier to a four channel electronic pulse analyser. (Cooke-Yarborough, Bradwell, Florida and Howells 1950.) The counter was placed in the horizontal gap of a large electromagnet. The pulses from the counter were fed through a long cable to the head amplifier which, with the rest of the electronic equipment, was removed as far as possible from the influence of stray magnetic fields. Several types of counter fillings were tried. To test their relative merits, a small quantity of A^{37} was admitted to the counter at each filling and the pulse distribution from it was measured. The width of the peak in the distribution is very sensitive to the presence of impurities. The percentage half width at half height of the A^{37} K capture peak was found to be 10·5 per cent by Hanna, Kirkwood and Pontecorvo. With argon and methane fillings, measured values of the half width were usually as high as 13–14 per cent, because the commercial methane used contained a considerable amount of impurities. Since methane cannot be purified very easily, carbon dioxide was tried instead. With argon and carbon dioxide fillings, values of the half width were reduced to 12–13 per cent; the carbon dioxide was purified by freezing, and pumping off any residual gas, repeating the process several times. When, in addition, the argon was purified by allowing it to circulate over heated calcium for several hours the half widths were reduced further in value to between 10·5 per cent, and 11·5 per cent. The best mixture was found to be about 10 parts of argon to 1 part of carbon dioxide; too great a quantity of carbon dioxide gives a greater half width, too small a quantity causes instability of the multiplication factor.

§ 3. EFFECT OF MAGNETIC FIELD ON THE COUNTER MULTIPLICATION PROPERTIES.

It was thought that the influence of the magnetic field on the motion of the avalanche electrons and the electrons produced by the ionizing particle might affect the multiplication properties of the counter, so the following experiment was performed in order to test this point. A small quantity of A^{37} was admitted into a counter, of 2 cm. diameter and 22 cm. active length, which was filled with a mixture of argon (21·8 cm. pressure) and methane (7·6 cm. pressure). The pulse distribution due to electrons emitted following the K capture process in A^{37} has a peak at an energy of 2·8 keV. The range of these electrons is very small, so that virtually none escape from the counter. Any change in the pulse distribution when the counter is placed in a magnetic field must therefore indicate that the multiplication properties of the counter have been affected in some way. Fig. 1 shows the pulse size distributions measured with the counter axis parallel and perpendicular to a magnetic field of 7500 gauss and without the field. The points all lie on the same curve indicating that in this case the magnetic field does not influence the counter properties. Later in the work the experiment was repeated with a counter having a larger diameter, and in this case a slight reduction in pulse size was observed in the presence of the field. Fig. 2 shows the distributions obtained with this counter which had a diameter of 5 cm. and an active length of 34 cm. It was filled with a mixture of argon (24·4 cm. pressure) and carbon dioxide (2·4 cm. pressure). The effect of the magnetic field is quite small and since the shape of the distribution is unaltered it is unimportant for many applications of the technique.

Fig. 1.



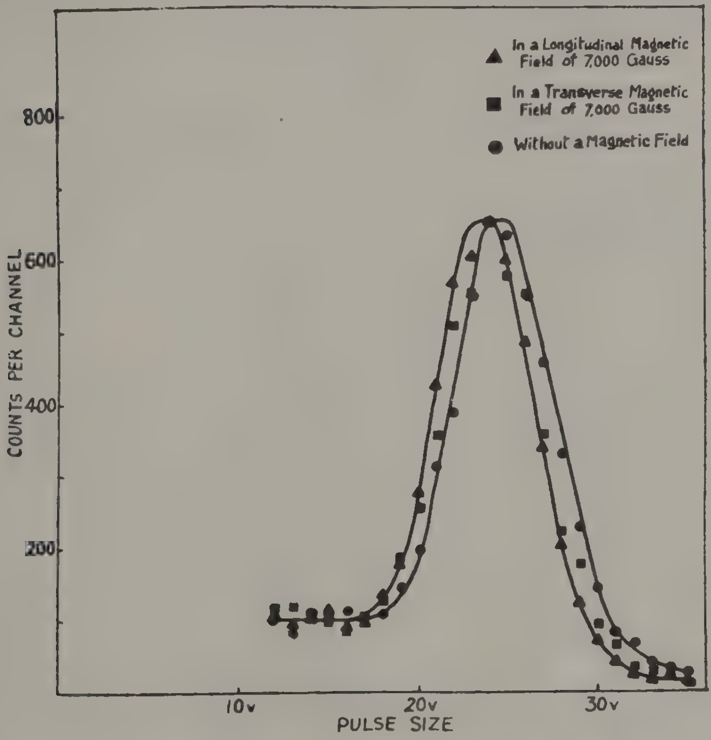
Effect of a magnetic field (7,500 gauss) on the pulse size distribution from A^{37} in the 2cm. diameter counter.

§ 4. Energy Calibration of the Counters.

Some means of calibrating the counters had to be devised before any measurements could be carried out. The usual method employs characteristic X-rays for this purpose. Since it is inconvenient to use an X-ray machine near a large magnet, long-lived radioactive isotopes which decay by K electron capture may be used as an alternative source of X-rays. A^{37} (2·8 keV.) has been used by Pontecorvo, Kirkwood and Hanna, and Fe^{55} (5·9 keV.) by Maeder but X-ray lines of higher energies than these are needed.

K-capture isotopes which can be prepared by thermal neutron irradiation, relatively free from long period β activities, were used. Specific activities of a few milli-curies per gm. gave adequate intensity. Table II. lists the K-capture activities which were obtained by irradiation in the Harwell pile.

Fig. 2.



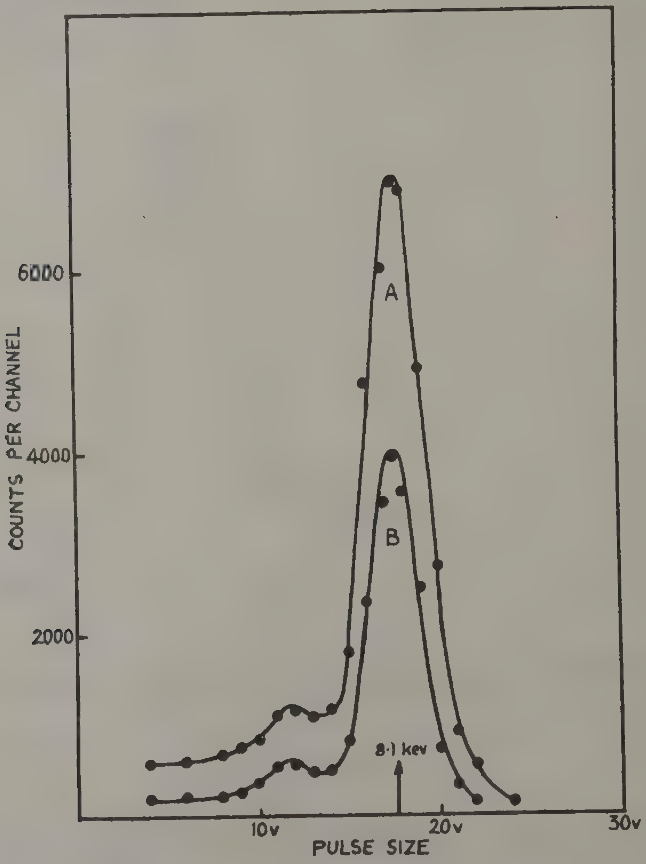
Effect of a magnetic field (7,000 gauss) on the pulse size distribution from A^{37} in the 5cm. diameter counter.

TABLE II.

Isotope	Half period (days)	Energy of Kα X-ray (keV.)
Cr^{51}	26·5	5·0
Zn^{65}	250	8·1
Se^{75}	127	10·5
Pd^{103}	17	20·2
Sn^{113}	105	24·1
Gd^{153}	155	41·6
Yb^{169}	33	50·8

With the lower energy X-rays from Cr, Zn and Se, it was essential to use thin sources and to provide a thin window in the counter wall to admit the radiation. With thick sources, the X-rays emerge only from a thin surface layer and are masked by gamma rays accompanying the K-capture, process, which can emerge from the body of the source. Thin sources were prepared by dissolving the irradiated element in acid and then evaporating a few drops of the solution on a mica foil. The sources were tested with a counter filled to a pressure of about 30 cm. of mercury with argon and carbon dioxide.

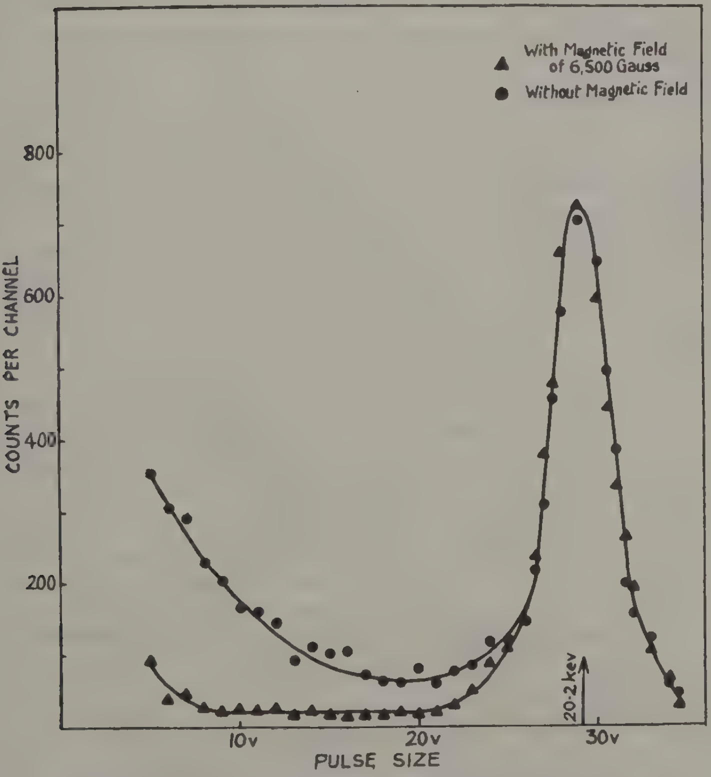
Fig. 3.



Pulse size distribution from Zn^{65} source.
Curve A: without critical absorber.
Curve B: with critical absorber (·0005″ Ni).

Fig. 3 curve A shows the pulse size distribution obtained with a Zn^{65} source (8·1 keV.). Curve B shows the distribution when a critical absorber (0·0005 in., nickel) which removes the Kβ radiation is placed between the source and the counter window. The Kβ radiation has an energy of 8·9 keV. and an initial intensity of about one eighth that of the Kα radiation. As expected the Kα and Kβ radiations are not resolved in curve A and the effect of the critical absorber is merely to reduce the width of the distribution slightly. The small peak on the low energy side of the main peak is due to the escape of the K radiation of argon (2·9 keV.) from the counter following the photoelectric absorption of the incident X-rays in the K shell of argon. Similar distributions having well defined peaks are obtained from the Cr^{51} and Se^{75} sources.

Fig. 4.

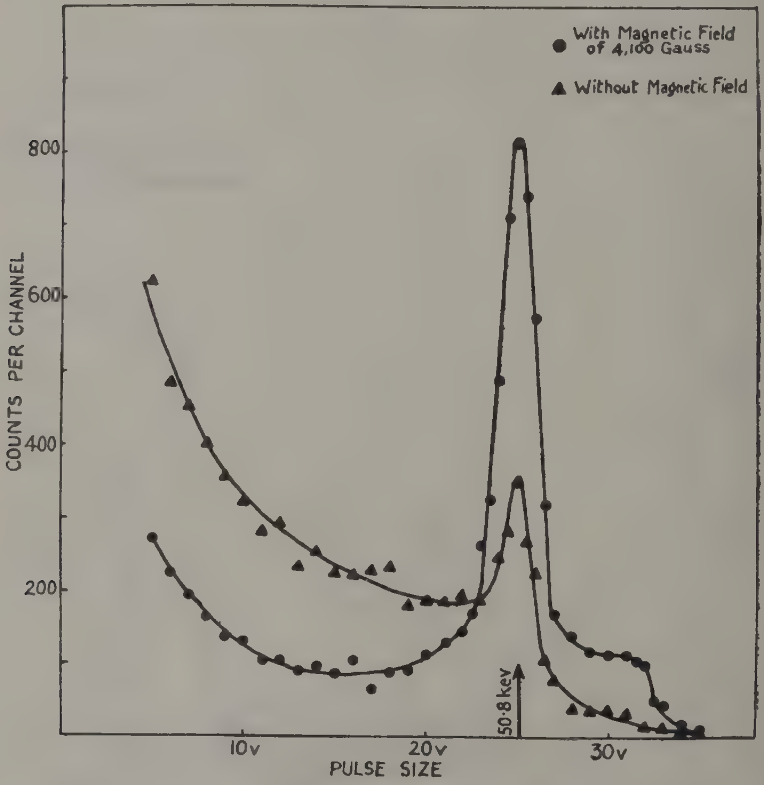


Pulse size distribution from Pd^{103} source with, and without, a magnetic field (6,500 gauss).

With the higher energy X-rays from Pd^{103}, Sn^{113}, Gd^{153} and Yb^{169}, it is not necessary to use very thin sources and the radiations are sufficiently hard to penetrate the glass wall of the counter (1 mm. thickness) without suffering too much attenuation.

Background radiations, consisting of β rays or electrons emitted by the source, or electrons generated in the walls by quanta, give rise to a continuous distribution of pulse sizes; these are greatly reduced by a magnetic field. An electron which originates in the cylindrical wall is rapidly bent back into the wall by the field and spends only a very small amount of energy in the counter. Fig. 4 shows the distribution from a Pd^{103} source (20·2 keV.) with and without a magnetic field and demonstrates this effect very clearly. A similar distribution is obtained from a Sn^{113} source. With the Gd^{153} and Yb^{169} sources, the photoelectrons generated by the X-rays are of sufficiently high energy for a significant fraction of them to escape from the counter at the pressure used. Fig. 5 shows the pulse distribution obtained from a Yb^{169} source (50·8 keV.) with and without a magnetic field. It can be seen that the magnetic field greatly increases the intensity in the peak, as well as decreasing the background intensity.

Fig. 5.



Pulse size distribution from Yb^{169} source with, and without, a magnetic field (4,100 gauss).

§5. ISOMERISM OF Br^{80}.

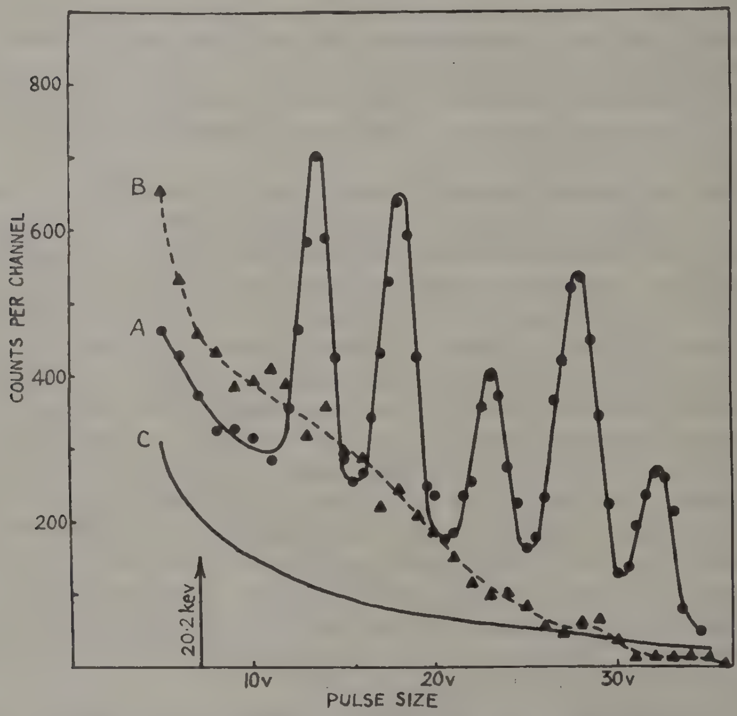
An investigation has been made of the well known case of isomerism in Br^{80}, by admitting a small quantity of carrier free methyl-bromide into a counter and measuring the pulse distribution from it. It is known that the 4·4 hr. isomer of Br^{80} decays into the lower state, an 18 min. β active body, with the emission of a 49 keV. γ-ray and a 37 keV. γ-ray in cascade (Grinberg and Roussinow 1940, Berthelot 1944). The 49 keV. γ-ray is totally internally converted, but there is still some doubt about the value of the conversion coefficient of the 37 keV. γ-ray and the type of transition. Berthelot has obtained a value of 0·64 for this conversion coefficient, by coincidence absorption measurements, and concludes that the transition is magnetic dipole. Segrè and Helmholtz (1949) suggest from parity considerations that it is more likely to be electric dipole, in which case the internal conversion coefficient should have a value different from that found by Berthelot.

The method used for producing the active methyl bromide was suggested by Dr. Glueckauf (Glueckauf, Jacobi and Kitt, 1949). A solution of potassium bromide in acetic acid was irradiated for a short period in the Harwell pile. Argon was bubbled through the solution, carrying active methyl bromide, generated during the irradiation, with it. The argon was then passed through a solution of caustic potash and then over phosphorus pentoxide to remove free bromine and water vapour. It then passed into a container cooled by liquid oxygen, where the methyl bromide was frozen down while the argon passed through. The container, still cooled by the liquid oxygen was transferred to the counter filling system and pumped free from air and other gases. The liquid oxygen was then removed and the methyl bromide allowed to evaporate into the counter.

Curves A and B in fig. 6 show the pulse size distribution from the active methyl bromide in an argon filled counter (containing about 30 cm. pressure of argon and carbon dioxide), with and without a magnetic field of 7000 gauss. This demonstrates very clearly the effect the magnetic field has in reducing electron escape.

The five peaks in the distribution arise in the following way. the 49 keV. gamma ray is totally internally converted, but the 37 gamma ray only partially, there are peaks in the distribution at $(37+49)=86$ keV., when both gamma rays are internally converted. and at 49 keV., when the 37 keV. gamma ray is unconverted and escapes from the counter. The chance of a 37 keV. γ-ray being absorbed in the counter is very small and can be neglected in our counter for pressures of argon less than 1 atmosphere. Both the peaks are modified by the escape from the counter of a 12 keV. Bromine K X-ray which is emitted in some cases during the filling of the vacancy created in the K shell of Bromine when a gamma ray is internally converted. Additional peaks will therefore be produced at $(49-12)=37$ keV. and at $(49-12+37)=(49+37-12)=74$ keV. when one Bromine K X-ray escapes, and at $(49-12+37-12)$ when two Bromine K X-rays escape. Peaks are therefore expected at 37, 49, 62, 74 and 86 keV. The counter was calibrated with 20·2 keV. X-rays from a Pd^{103} source, the position of the calibration peak is indicated in fig. 6 and it can be seen that the five peaks occur at about the expected energies. Fig. 7 shows the pulse size distribution from the active methyl bromide in the same counter filled with krypton at a pressure of 66·5 cm. and carbon dioxide (4 cm.). The escape probability of the bromine K X-ray is 99 per cent in the argon counter, but only 45 per cent in the krypton counter; it can be seen that the relative intensities of all the "K escapes" peaks have been reduced in the distribution obtained with krypton. Conversion coefficients can be calculated from these curves. The 49 keV. γ-ray is totally converted, which is verified by the absence of a peak at $(37-12)=25$ keV., so the conversion coefficient of the 37 keV. γ-ray is given by

Fig. 6.



Pulse size distribution from Br^{80} in an argon filled counter.
Curve A: with counter in a magnetic field of 7,000 gauss.
Curve B: without magnetic field.
Curve C: background from Br^{82} (corrected for decay) with counter in a magnetic field of 7,000 gauss.

$$\alpha_{37}=\frac{\text{no. of converted 37 keV. }\gamma\text{-rays}}{\text{no. of unconverted 37 keV. }\gamma\text{-rays}}=\frac{I_{62}+I_{74}+I_{86}}{I_{49}+I_{37}},$$

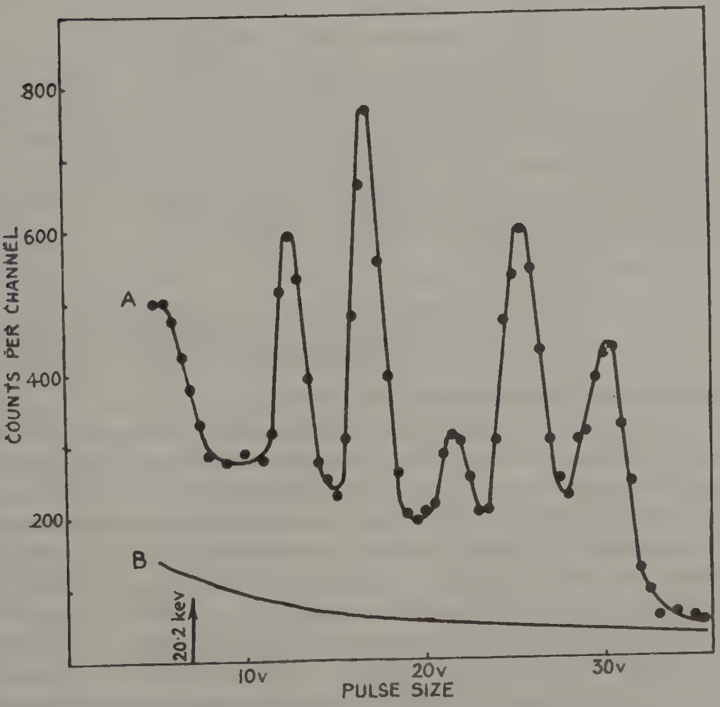
where I_{62} represents intensity in the 62 keV. (peak), etc.

Table III. gives values of the conversion coefficient obtained in this way for various runs with several different counter fillings and bromine samples. In the case of the measurement with krypton a correction must be applied for reabsorption in the counter of the 37 keV. γ-ray. The probability of escape is calculated to be 0·92 for a 37 keV. γ-ray in the krypton filled counter.

TABLE III.

Counter filling	Conversion coefficient
$A+CO_2$	1·3
$A+CO_2$	1·1
$A+CO_2$	1·2
$A+CO_2$	1·4
$Kr+CO_2$	1·2
	av. 1·2

Fig. 7.



Pulse size distribution from Br^{80} in a krypton filled counter.
Curve A: distribution in a magnetic field of 7000 gauss.
Curve B: background distribution from Br^{82} (corrected for decay) in a magnetic field of 7000 gauss.

The main error in these determinations arises from the background, composed partly of 34 hour Br^{82}, which can be measured when the 4·4 hour activity has decayed, and partly of 18 min. Br^{80} in equilibrium with the 4·4 hour Br^{80} activity, which cannot be separated from it. Curve C, fig. 6, represents the 34 hour background, corrected for decay. The conversion coefficients were evaluated

(1) taking a curve close to C as the true background;
(2) taking a curve drawn close to the troughs between the peaks as the background.

There is a difference of about 14 per cent between these two extreme values of the measured conversion coefficient, and it is the mean which has been tabulated in each case in Table III.

The value of the conversion coefficient calculated for (*a*) a magnetic dipole transition, (*b*) an electric dipole transition, is 0·60 and 1·55 respectively. (Hebb and Nelson 1940). Our value of 1·2 for the conversion coefficient proves that the transition is electric dipole rather than magnetic dipole. This conclusion is in contradiction with the results of Berthelot but agrees with a result of Lidofsky, Macklin and Wu (1950) which has just been briefly reported.

§ 6. FLUORESCENCE YIELD OF KRYPTON.

There are two different ways in which a vacancy in an inner shell of the electron structure of an atom may be filled; either a characteristic X-ray is emitted when an electron falls in from a higher level or the corresponding energy is shared between several electrons (Auger electrons), ejected from higher electronic levels. The fluorescence yield is defined for a vacancy in a particular shell of the atom in the following way.

$$\text{Fluorescent yield} = W_{K,(LM,\ldots.)} = \frac{\text{No. of K, (L,M . . . .) X-ray photons emitted}}{\text{No. of vacancies in K, (L,M . . . .) shell.}}.$$

Fluorescence yields were first measured by Auger in 1926. He investigated gaseous materials in a cloud chamber using X-rays to excite the atoms. When the incident X-ray is absorbed in the K shell a photoelectron of energy $E_0 - E_K$ (where E_0=energy of incident X-rays and E_K=binding energy of the K shell of the absorbing atom) is observed. It is accompanied in W_K cases, *i. e.* when a K X-ray is emitted, only by several low energy Auger electrons from vacancies in the L,M . . . . shells. (The fluorescence yield is much smaller for the L,M . . . . shells than for the K shell so that the probability of emission of one or more Auger electrons during the cascade process of filling a vacancy in the L,M . . . . shells is very high). In the remaining $(1—W_K)$ cases a moderately energetic electron of energy $X_K—E_{L,M}$ . . . . (X_K=energy of K radiation, $E_{L,M}$ . . . . =binding energy of L.M . . . . . shells) is observed as well. The fluorescence yield was obtained from the ratio of the number of events of each type.

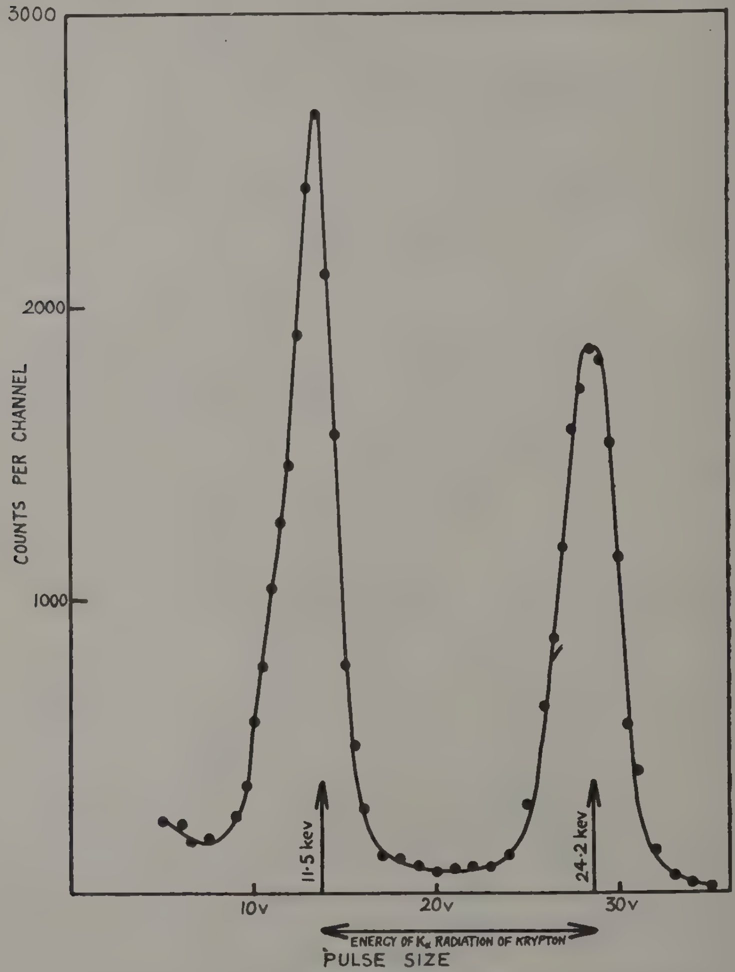
A method for measuring fluorescence yields of solid materials was devised by Compton (1929). An ionization chamber is used to measure the number of X-rays (whose energy must be greater than the K binding energy of the element being examined) absorbed in a block of the material; the intensity of the fluorescent X-rays excited in the material is measured at the same time. After correcting for absorption of the incident X-rays in the L,M . . . . shells the fluorescence yield is obtained directly.

The proportional counter, as pointed out by Kirkwood, Pontecorvo and Hanna, provides an alternative method of measuring the fluorescence yields of gases. When an X-ray is absorbed by the K shell of the counter gas, and the vacancy caused by the emission of the photo-electron is filled with emission of Auger electrons, the full energy of the incident X-rays is spent in the counter. When the vacancy is filled with emission of a K X-ray, however, there is a good chance that this X-ray will escape from the counter, provided the pressure in the counter is low. In this case an energy E_0—X_K (where E_0 is the energy of the incident X-rays and X_K is the energy of the *K radiation*) is spent in the counter. Thus, two peaks are observed in the pulse size distribution and the fluorescence yield can be determined from their relative intensities.

Fig. 8 shows a typical pulse size distribution obtained by irradiating a counter containing krypton and carbon dioxide (16 parts krypton: 1 part carbon dioxide) with characteristic X-rays (24·2 keV.) from a Sn^{113} K capture source. A magnetic field was applied to reduce the escape of photo-electrons from the counter and to reduce background radiations from the source. The energy of the K_α X-ray of krypton is 12·7 keV. so when this radiation escapes a peak is obtained at $24{\cdot}2-12{\cdot}7=11{\cdot}5$ keV. An unresolved peak due to the escape of the K_β radiation of Kr (14·1 keV.) is also apparent. The K_β radiation from the source was removed by a critical absorber (0·005 silver) placed between the source and the counter.

Two corrections must be applied in order to evaluate the fluorescence yield, W_K, from the relative intensity of these two peaks. First, the intensity of the "escape" peak is reduced by reabsorption of the K X-rays of krypton in the counter gas. The probability of reabsorption was computed from the known absorption coefficients using unpublished calculations of B. Pontecorvo. This correction can be made small by using low pressures of gas and did not exceed 5 per cent in any of the measurements. A second correction was necessary to take account of the incident radiation absorbed in the L,M . . . . shells. Since the total energy of the incident X-ray is then spent in the counter, this process cannot be distinguished from K absorption without subsequent emission of a K X-ray. The ratio of the absorption in the K shell to that in the L,M . . . . shells is known at the K absorption edge for any element and it is assumed that the ratio is the same for energies not much greater than the K binding energy. The ratio for Kr was taken to be 6·4. This value had to be calculated from an empirical formula due to Jönsson (see Compton and Allison) since there are very few measurements of the X-ray absorption coefficient of krypton reported in the literature. It follows, therefore, that absorption in the L,M . . . . shells constitutes a fraction 1/7·4 of the total absorption in krypton. A similar correction had to be applied in the technique used with solid materials mentioned above. The possiblity of an error due to escape of energetic photo-electrons from the counter was checked by measurements taken with magnetic fields of 3000, 4600, and 5900 gauss with a Sn^{113} source. The relative intensities of the escape peak were 0·55, 0·56 and 0·56 respectively, which shows that the escape is small.

Fig. 8.



Pulse size distribution from Sn^{113} source (24·2 keV) in a krypton filled counter (magnetic field 4300 gauss).

Absorption of X-rays in the CO_2 was neglected since it is very small compared with that in krypton.

The areas under the " escape " and " normal " peaks were measured by planimeter from graphs of the pulse size distribution. The total number of counts recorded in each peak exceeded 6000 counts in every case. The results obtained with different pressures and using Pd^{103} (20·2 keV.) and Sn^{113} (24·2 keV.) sources are given in Table IV.

The mean value of the fluorescence yield from these results is 0·67. The probable error of the average is only 0·005, but the systematic error, which is mostly due to the L absorption correction, is certainly larger than this.

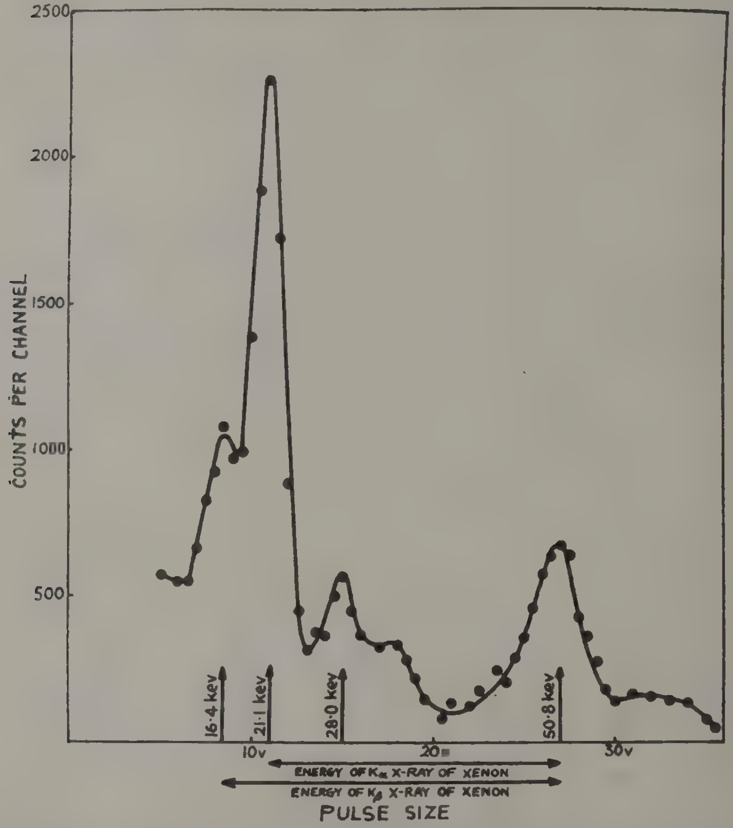
Values given in the literature for the fluorescence yield of krypton are 0·51 (Auger 1926) and 0·615 (Harms 1927). Locher (1932) gives a value of 0·585 obtained by interpolation from data for neighbouring elements. The value obtained by the proportional counter technique is therefore considerably higher than that obtained by previous measurements.

TABLE IV.

Energy of incident X-rays (keV.)	Magnetic field (Gauss)	Pressure of krypton (cm. of mercury)	Fraction of pulses in "escape" peak	Fraction of pulses in "escape" peak		Mean fluorescence yield
				Corrected for re-absorption of K radiation	And corrected for L,M... absorption	
24·2	3,800	5·4	0·59	0·60	0·70	0·66
24·2	3,000	7·4	0·55	0·56	0·65	
24·2	4,600	7·4	0·56	0·57	0·66	
24·2	5,900	7·4	0·56	0·57	0·66	
24·2	4,400	13·8	0·52	0·55	0·63	
20·2	3,800	2·9	0·59	0·60	0·69	0·68
20·2	7,100	2·9	0·60	0·60	0·70	
20·2	3,800	5·4	0·57	0·59	0·68	
20·2	3,800	7·4	0·57	0·59	0·68	
20·2	4,400	13·8	0·55	0·58	0·67	

A measurement of the fluorescences yield of xenon was also attempted, but in this case the accuracy is greatly limited by background radiations from the K capture source used. Fig. 9 shows the distribution from a Yb^{169} source (50·8 keV.) in a xenon filled counter. The main "escape" peak is at 50·8—29·7=21·1 keV. (since the energy of the K_α ray of xenon is 29·7 keV.), and is much more intense than the full energy peak. The subsidiary peak at 16·4 keV. is due to escape of the K_β radiation of xenon (34·4 keV.). This was shown to be resolved when the escape peak was investigated in greater detail than in fig. 9. The K_β radiation from the source, of energy 57·7 keV. cannot be distinguished in the main peak, but is easily seen in the "escape" peak, at 57·7—29·7=28·0 keV. This suggests that escape peaks may be useful in investigating radiations whose total energies are too close to be resolved directly. The remaining peak probably arises from some impurities in the Yb. The value of the fluorescence yield of xenon obtained was 0·81. In view of the background and the fact that only a single measurement was taken, the accuracy is much smaller than in the case of krypton.

Fig. 9.



Pulse size distribution from Yb^{169} (50·8 keV) in a xenon filled counter (magnetic field 7,000 gauss).

Acknowledgments.

We are grateful to Dr. B. Pontecorvo for his valuable advice and help in this work, and to Drs. W. D. Allen and R. H. V. M. Dawton who kindly allowed us to use a large electromagnet. Thanks are also due to Mr. F. Bradley for his assistance on many occasions.

The work was done in the Nuclear Physics Division at Harwell and acknowledgment is made to the Director, A. E. R. E., Harwell, for permission to publish these results.

REFERENCES

ANGUS, J., COCKROFT, A. L., and CURRAN, S.C., 1949, *Phil. Mag.*, **40**, 522.

AUGER, P., 1926, *Ann. Phys., Paris*, **6**, 183.

BERTHELOT, A., 1944, *Ann. Phys., Paris*, **19**, 219.

COMPTON, A. H., 1929, *Phil. Mag.*, **8**, 961.

COMPTON, A. H., and ALLISON, S.K., 1935, *X-rays in Theory and Experiment* (New York: D. Van Nostrand Co., Inc.).

COOKE-YARBOROUGH, E. H., BRADWELL, J., FLORIDA, C. D., and HOWELLS, G. A., 1950, *Proc. Instn. Elect. Engrs.*, **97**, 108.

CURRAN, S. C., ANGUS, J., and Cockroft, A. L., 1948, *Nature, Lond.*, **162**, 302; 1949, *Phil. Mag.* **40**, 36, 53, 305, 929.

GLUECKAUF, E., JACOBI, R. B., KITT, G. P., 1949, J. *Chem. Soc.*, S.330.

GRINBERG, A. P., and ROUSSINOW, L. I., 1940, *Phys. Rev.*, **58**, 181.

HARMS, M. I., 1927, *Ann. d. Physik*, **82**, 87.

HANNA, G. C., KIRKWOOD, D. H. W., and PONTECORVO, B., 1949, *Phys. Rev.*, **75**, 985.

HANNA, G. C., PONTECORVO, B., 1949, *Phys. Rev.*, **75**, 983.

HEBB, M. H., and NELSON, E., 1940, *Phys. Rev.*, **58**, 486.

KIRKWOOD, D. H. W., PONTECORVO, B., HANNA, G. C., 1948, *Phys. Rev.*, **74**, 497.

LIDOFSKY, L., MACKLIN, P., and WU, C. S., 1950, *Bull. Am. Phys. Soc.*, **25**, 10.

LOCHER, G. L., 1932, *Phys. Rev.*, **40**, 484.

MAEDER, D., Private communciation.

MAZE, R., 1946, *J. Phys. Radium*, **6**, 164.

PONTECORVO, B., KIRKWOOD, D. H. W., HANNA, G. C., 1949, *Phys. Rev.*, **75**, 982.

ROTHWELL, P., and WEST, D., 1950, *Proc. Phys. Soc. A*, **63**, 539, 541.

SEGRÈ, E., and HELMHOLZ, A. C., 1949, *Rev. Mod. Phys.*, **21**, 271.

LXXXI. *The Eddy Viscosity in Turbulent Shear Flow.*

By A. A. TOWNSEND,
Emmanuel College, Cambridge *.

[Received May 26, 1950].

SUMMARY.

The comparative slowness of those processes in the turbulent wake which determine the turbulent energy balance makes it necessary to consider the whole wake flow in attempts to derive local values of such quantities as turbulent intensity and eddy viscosity. If the wake is in a state of moving equilibrium, *i. e.* the development involves processes controlling the intensity and eddy viscosity to values appropriate to the particular stage of development, then the controlling mechanism must be of large scale both in space and time. It is suggested that the large slow eddies responsible for the phenomenon of intermittently turbulent flow form this mechanism, and a plausible control sequence is outlined in which the intensity of the large eddies controls the level of eddy viscosity. The detailed structure of these large eddies has been investigated by measuring the spectrum functions of the velocity fluctuation components in the wake of a circular cylinder. The group of large eddies is distinct, and it is possible to deduce that the large eddies are randomly placed in the mean stream direction, extend over most of one half of the wake, and have their vorticity roughly in the direction of maximum positive mean rate-of-strain. By a rough calculation it is shown that this type are near the state of energy equilibrium necessary for the operation of the suggested control process. Finally, the generation of Reynolds shear stresses by distortion of quasi-stationary turbulent fields is discussed, and possible reasons for the applicability of the eddy viscosity hypothesis are suggested.

§ 1. INTRODUCTION.

RECENT work by the author on the structure of the turbulent wake of a cylinder has shown that wake development is due to the existence of a group of large-scale eddies whose function is mainly convective (Townsend 1949 and 1950). That is, they convect fully turbulent fluid from near the wake centre to the outer part of the wake, and simultaneously entrain previously undisturbed fluid bringing it into closer contact with the fully turbulent fluid of the wake core. The appearance of intermittently turbulent flow in the wake is a natural consequence of this process, and is the most striking manifestation of the presence of these eddies which contain only a relatively small fraction of the total turbulent energy. The complete

---

* Communicated by the Author.

turbulent motion can be regarded as the superposition of a small-scale random motion containing most of the turbulent energy on this large-scale convective motion. Detailed measurements show that the intensity, scale and diffusive properties of the fully turbulent fluid are substantially constant at any section of the wake, irrespective of distance from the wake centre, and this is held to be a natural consequence of the mixing action of the large eddies, which causes the instantaneous properties of a sample of turbulent fluid to depend more on its past history of crossing and recrossing the wake than on its present position. This notion that the properties of the turbulent fluid are not directly controlled by the instantaneous shear is contrary to the usual assumptions of mixing-length theory, which are based on the notion of local and instantaneous equilibrium of the turbulent intensity. If the statistical equilibrium of the wake properties is controlled more by the past history of the wake than by its present state, it is not possible to use this notion to determine the equilibrium turbulent intensity, and this paper describes an alternative approach, based on a consideration of the dynamic equilibrium of the large eddies.

§ 2. NOTATION FOR THE CYLINDER WAKE.

The usual notation for turbulent flow with two-dimensional mean motion will be used, that is, the components of the mean velocity along the axes Ox, Oy, Oz are U, V, 0, and the instantaneous components of the fluid velocity are $U+u$, $V+v$, w. Fluid density is denoted by ρ, kinematic viscosity by ν, and the instantaneous pressure by $P+p$, where p is the mean pressure. All mean values are taken with respect to time at a point fixed in space.

For the cylinder wake, the coordinate system is such that Oz is along the axis of the cylinder, Ox is parallel to the direction of the mean stream, and Oy is perpendicular both to the direction of the mean stream and to the axis of the cylinder. Then, making the usual "boundary layer" approximations (Goldstein 1938), the exact equations of mean motion reduce to one

$$U\frac{\partial U}{\partial x}+\frac{\partial \overline{uv}}{\partial y}=\nu\frac{\partial^2 U}{\partial y^2}. \qquad (1)$$

If it is assumed that the turbulent shear stress $\overline{uv}$ can be described in terms of a coefficient of eddy viscosity ϵ, defined by

$$\overline{uv}=-\epsilon\frac{\partial U}{\partial y}$$

then

$$U\frac{\partial U}{\partial x}=\frac{\partial}{\partial y}\left[(\nu+\epsilon)\frac{\partial U}{\partial y}\right]. \qquad (2)$$

Using the same assumptions as in (1), the equation for the balance of turbulent energy is found to be

$$\overline{W}+\tfrac{1}{2}U_0\frac{\partial}{\partial x}(\overline{u^2}+\overline{v^2}+\overline{w^2})+\overline{uv}\frac{\partial U}{\partial y}+\frac{1}{2}\frac{\partial}{\partial y}(\overline{u^2v}+\overline{v^3}+\overline{vw^2})+\frac{1}{\rho}\frac{\partial \overline{pv}}{\partial y}=0, \qquad (3)$$

where $\overline{W}$ is the mean viscous dissipation

$$\overline{W}=\nu(\overline{u\nabla^2 u}+\overline{v\nabla^2 v}+\overline{w\nabla^2 w}).$$

If the flow is at a sufficiently high Reynolds number to permit the application of the theory of local similarity (Townsend 1948), then

$$\overline{W}=15\nu\overline{\left(\frac{\partial u}{\partial x}\right)^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and all the terms in equation (3) can be measured directly except $(1/\rho)$ $(\partial\overline{pv}/\partial y)$, which can be obtained by difference.

§ 3. THE MIXING-LENGTH HYPOTHESIS OF LOCAL ENERGY EQUILIBRIUM.

The equation for the turbulent energy balance (equation (3)) contains terms representing viscous dissipation, convection of energy by the mean flow, production of turbulent energy by working against the turbulent shear stress, energy transport by diffusive movements, and energy transport by flow down pressure gradients, and states that the energy density is constant in time as a result of a balance between the gains and losses of energy from these processes. By assuming that the turbulence is so small in scale that the equilibrium intensity is determined only by the local parameters of the flow, and further, that the motion is dynamically similar everywhere, it is possible to derive the Prandtl relation between turbulent intensity and rate of shear,

$$v'=Kl_1\left|\frac{\partial U}{\partial y}\right|. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

This assumption of small scale implies that the terms representing convection of energy by the mean flow, energy transport by diffusive movements and energy transport by pressure flow are negligible compared with those representing production of energy by shear and loss by viscous dissipation. If this assumption is not made, the energy equation produces no definite results and the simplicity of the attack is lost. Several attempts have been made to allow for the effects of convection and diffusion, but it has been shown experimentally that all the terms in the energy equation are of comparable magnitude, and that it is very unlikely that the mixing-length view of the energy balance is even a good approximation. This is most clearly shown by considering the time necessary for a chance fluctuation in turbulent intensity to disappear. Since the local rate of production of turbulent energy is $-\overline{uv}(\partial U/\partial y)$ and the local energy density is $\frac{1}{2}(\overline{u^2}+\overline{v^2}+\overline{w^2})$ this time must be of order

$$-\tfrac{1}{2}(\overline{w^2}+\overline{v^2}+\overline{w^2})\Big/\overline{uv}\frac{\partial U}{\partial y},$$

which experimentally is found to have a minimum value of $1{\cdot}5\,x/U_0$, *i. e.* more than the time necessary for the free stream to pass from the cylinder to the point of observation (Townsend 1950). This estimate of the "relaxation time" is supported by the experimental observation that the setting up of dynamical similarity in wake flow is a very slow process, and is not complete even at a distance of 500 diameters downstream from the cylinder.

With the knowledge that diffusive and pressure transport of energy are sufficiently large to change the energy balance by a considerable amount, it can only be concluded that energy equilibrium in the wake is a process involving the whole wake, and that the turbulent intensity in a sample of turbulent fluid is determined more by its past history of crossing and re-crossing the wake than by its present position. This conclusion, while it explains very naturally the observed invariance of the mean properties of the turbulent fluid across any wake section as a consequence of the statistically uniform history of all samples of fully turbulent fluid, makes the problem of computing the actual turbulent intensity from the energy equation almost impossible and a new approach is necessary.

§ 4. THE HYPOTHESIS OF MOVING EQUILIBRIUM IN WAKE FLOW.

While the kind of absolute local equilibrium considered in mixing-length theory does not exist in free turbulence, a moving equilibrium may be possible, *i. e.* when the wake is in a steady state, the mean rate of production of turbulent energy over a whole section is controlled by the turbulent intensity in such a way that there is a stable equilibrium intensity for any wake section and that chance fluctuations from this intensity distribution will not persist as the wake develops. Considering the history of a small sample of turbulent fluid, it is possible that, when the turbulent intensity is above normal, increased loss of energy by viscous dissipation may outweigh the consequent rate of conversion of mean flow energy into turbulent energy and so tend to the equilibrium intensity sufficiently far down-stream. This may occur as the result of local changes in structure, similar to those implied in the Prandtl equation, or by overall changes caused by the increased turbulent intensity. The homogeneity of the fully turbulent fluid makes it very unlikely that local changes in structure control the equilibrium, but the group of large eddies could form a very effective large-scale controlling mechanism in spite of their comparatively small energy. These large eddies, conveying turbulent fluid from the wake centre into close contact with non-turbulent fluid from outside the wake, are primarily responsible for the lateral spreading of the wake and consequently control the rate of conversion of mean flow energy into turbulent energy. Taking the extreme examples, if they were to stop completely wake development would almost cease and the overall rate of conversion of mean flow energy to turbulent energy would become very small, while a large increase in their intensity would accelerate wake spread and increase the rate of turbulent energy production. This is easily shown by writing down the kinetic energy of the mean flow

$$\int_{-\infty}^{\infty} \tfrac{1}{2}\rho(\mathrm{U}_0-\mathrm{U})^2\,dy=\tfrac{1}{2}\rho\mathrm{U}_0^2\frac{d^2}{l},$$

since

$$\int_{-\infty}^{\infty} \rho\mathrm{U}_0(\mathrm{U}_0-\mathrm{U})\,dy=\tfrac{1}{2}\rho\mathrm{U}_0^2 d\mathrm{C}_d,$$

where l is a parameter representing the wake width, d is the cylinder diameter, C_d is the drag coefficient of the cylinder.

Then the total rate of production of turbulent energy over the whole wake section is

$$-U_0\frac{\partial}{\partial x}\int_{-\infty}^{\infty}\tfrac{1}{2}\rho(U_0-U)^2\,dy=\tfrac{1}{2}\rho U_0^3\frac{d^2}{l^2}\frac{dl}{dx},$$

proportional to the rate of increase of wake width.

§ 5. ENERGY BALANCE OF THE LARGE EDDIES.

If the large eddies control the rate of production of turbulent energy, it follows that their equilibrium intensity is itself controlled, and it is simpler to consider the factors in this control process than the factors controlling the intensity of the fully turbulent fluid. It is clear that the large eddies must obtain energy directly from the mean flow, and it seems likely that the loss of energy by friction of such a large scale motion may be described by the same coefficient of eddy viscosity as is effective for the mean flow. Then the control process should lead to a balance in which the large eddies are maintained in a condition of neutral equilbrium, the viscous dissipation only exceeding the energy gain from the mean flow by a sufficient amount to keep the wake in a condition of similarity. Any increase in the intensity of the large eddies leads to a too rapid spread of the wake, to increased production of turbulent energy, and finally to an increased eddy viscosity which decreases the intensity. If this is true, then for any particular mean velocity profile, the equilibrium condition will be of the form

$$\frac{u_1 l_1}{\epsilon}=R_c',$$

where R_c' is a pure number, characteristic of the shape of the mean velocity profile, u_1, l_1 are the velocity and length scales at the particular wake section.

Applying the momentum condition for the cylinder wake, this becomes

$$\frac{U_0 d}{\epsilon}=R_c.$$

This implies that the mean motion in a turbulent wake is always taking place at the effective Reynolds number R_c, whatever the cylinder Reynolds number, $R_d=U_0 d/\nu$, may be, which is one way of regarding the observed invariance with cylinder Reynolds number of the properties of the mean flow.

Experience with the laminar instability of a boundary layer suggests that the instability represented by the large eddies might be two-dimensional and periodic in space, but the experiments described below show that the large eddies are certainly three-dimensional and aperiodic. The implications of this superficially surprising results will be considered later.

§ 6. EXPERIMENTAL ARRANGEMENTS.

The measurements were all made in the wake of a circular cylinder of diameter 0·159 cm. placed in an air-stream of velocity 1280 cm. sec.$^{-1}$ and usually at a position 500 cylinder diameters downstream. The air-stream was produced in the small wind-tunnel in the Cavendish Laboratory, of section 38 cm. × 38 cm.. At the position of observation, the wake flow is very nearly in a state of similarity, and it is believed that the results obtained are substantially independent of the initial flow conditions close to the cylinder (Townsend 1950).

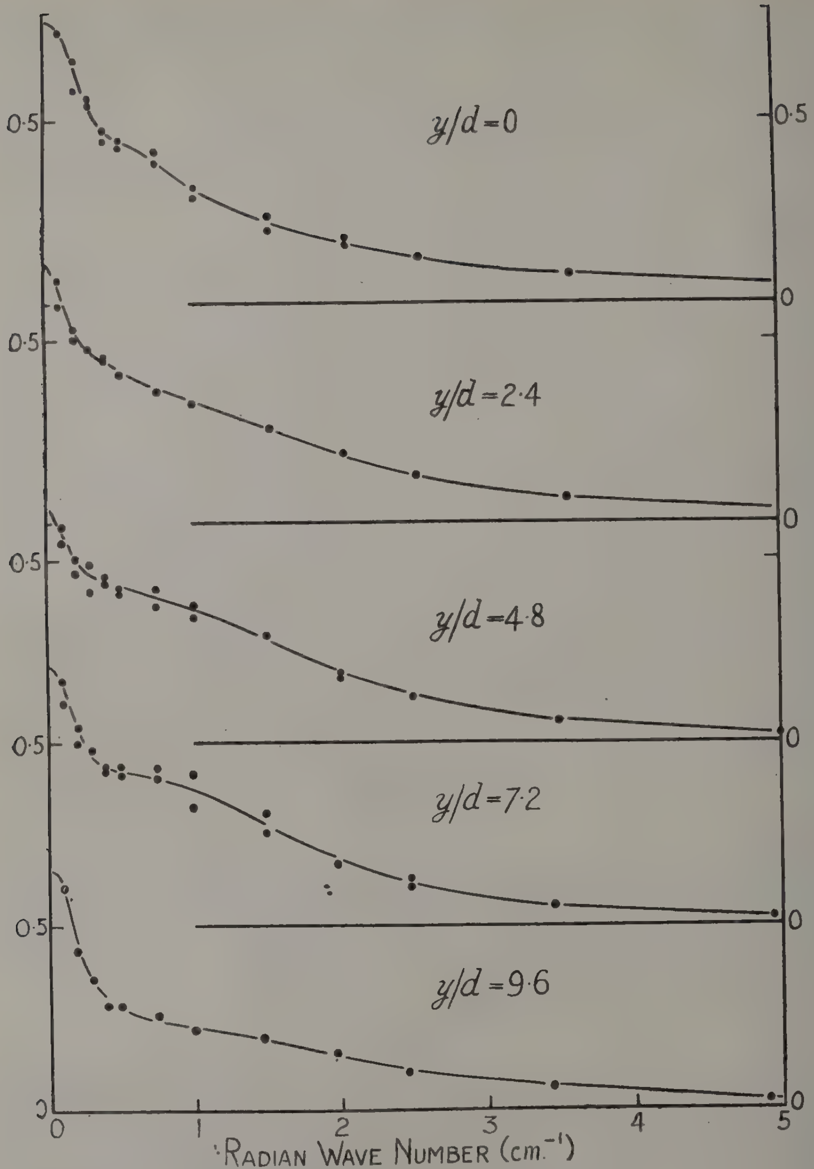
Turbulent velocity fluctuations were observed with the hot-wire anemometer, in both the single-wire and X-wire forms. Since a straight hot-wire responds very nearly to the component of the velocity fluctuation at right angles to its length and in the plane of the hot-wire and the mean flow direction, a single hot-wire at right angles to the mean stream registers fluctuations of u, one in the direction $(1/\sqrt{2}, -1/\sqrt{2}, 0)$ registers fluctuations in $(u+v)/\sqrt{2}$, and one in the direction $(1/\sqrt{2}, 1/\sqrt{2}, 0)$ fluctuations in $(u-v)/\sqrt{2}$. In this way the fluctuations components in the directions of the three axes, i. e. u, v, w, and also the components $(u+v)/\sqrt{2}$ and $(u-v)/\sqrt{2}$ could be detected and their intensities measured.

To obtain the intensity spectrum of a fluctuating electric signal, several methods are available. For simplicity, a zero-beat heterodyne analyser was used, a circuit in which, by means of a balanced modulator, the amplified turbulence signal is multipled by a sinusoidal signal from a beat-frequency oscillator, and the intensity of the resultant measured after passage through a low-pass filter. Since only the sum and difference frequencies appear in the resultant, the effect is that the original turbulence spectrum is displaced along the frequency axis by the oscillator frequency, and the energy transmitted through the low-pass filter is proportional to the energy originally contained in a band of frequencies, of central frequency the oscillator frequency and of band-width twice the cut-off frequency of the filter. This form of spectrum analyser is believed to have advantages for the study of the low-frequency part of the spectrum.

§ 7. PROPERTIES AND STRUCTURE OF THE LARGE SLOW EDDIES.

From previous work on the structure of the turbulent wake it is known that the large eddies, which are responsible for producing the intermittently turbulent flow in the cylinder wake, contain only about 5 per cent of the total turbulent energy and extend over the whole width of the wake. The spatial extent of the circulation in the down-stream direction appears to be even greater. Using circuits designed to measure intermittency factors directly, it has been shown that the occurrence of fully turbulent flow in one half of the wake is uncorrelated with its occurrence in the outer half (Townsend 1950), and it is likely that the eddies in the two halves are substantially independent. This result makes it unlikely that the large eddies represent a periodic instability similar to the Karman vortex street, but more definite evidence is found in spectrum measurements (figs. 1–4).

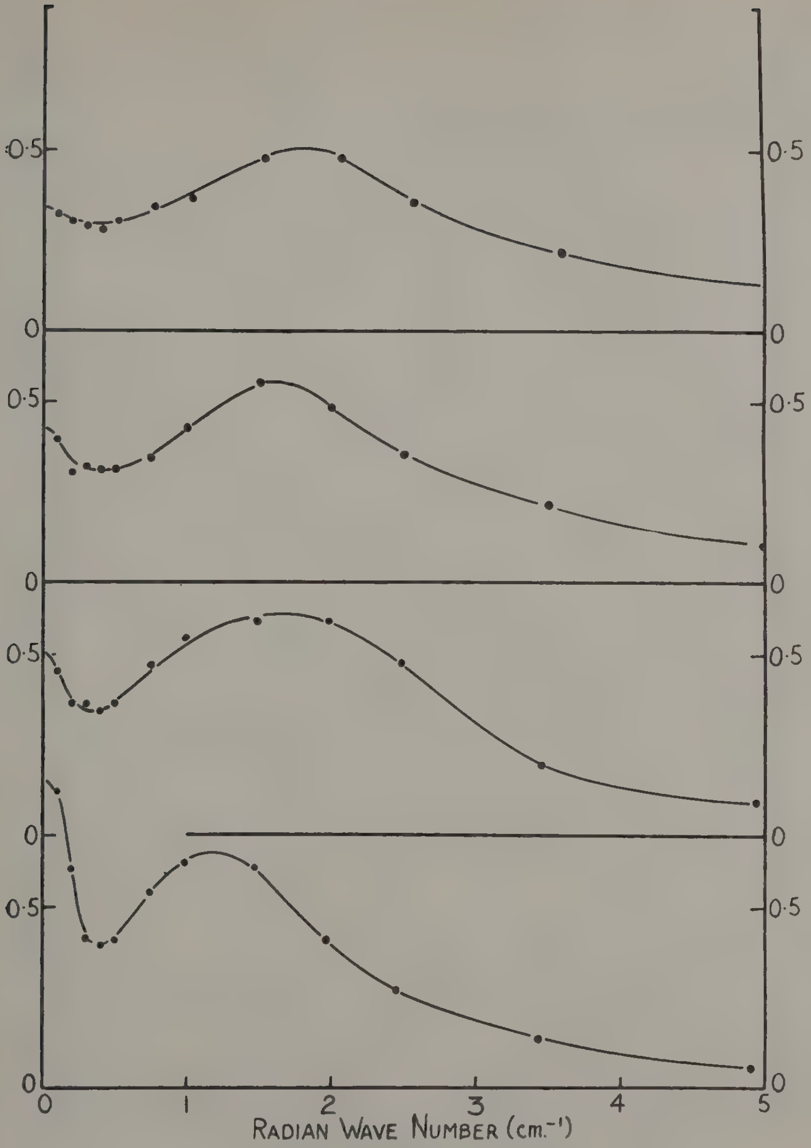
Fig. 1.



Spectrum functions of *u*-component.

The spectrum of the *u*-component of the velocity fluctuation has a characteristic form, clearly due to the superposition of two groups of turbulence of widely different scales (compare, for example, the spectrum of isotropic turbulence (fig. 5) which may be regarded as the spectrum of a

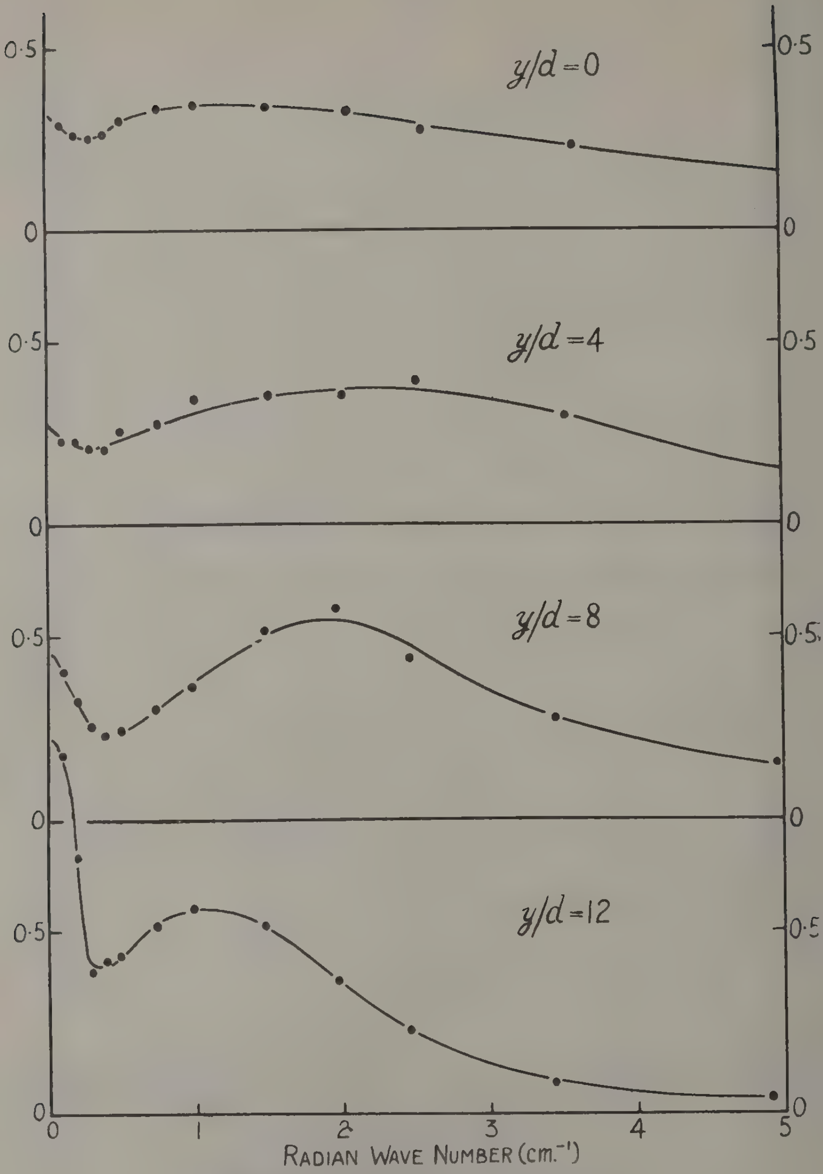
Fig. 2.



Spectrum functions of v-component.

random collection of eddies with a continuous distribution of size), and it is natural to identity the low frequency component of low total energy with the group of large slow eddies previously postulated to explain the experimental observation on diffusion and intermittency. Within the rather large experimental error, the spectrum function of the large slow eddies is a

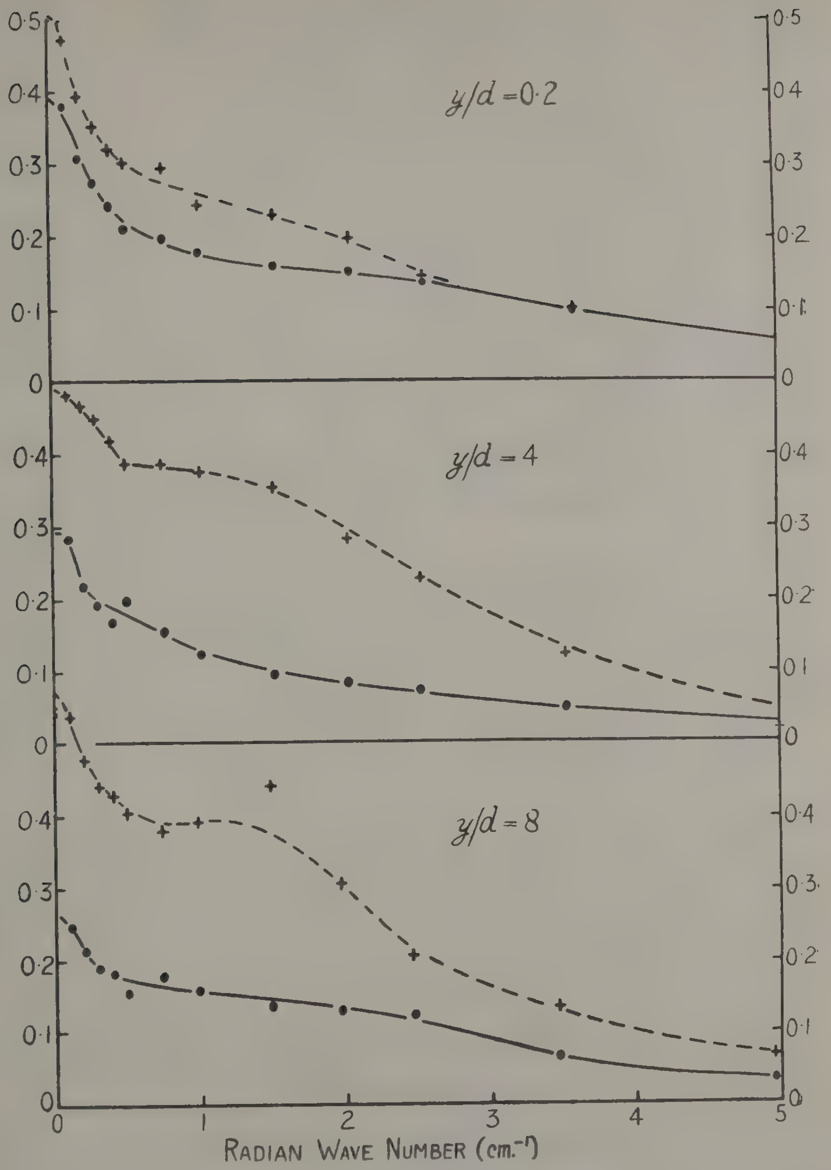
Fig. 3.



Spectrum functions of w-component.

Gaussian error function which is characteristic of an aggregate of similar eddies of simple structure, uncorrelated in position along the Ox axis, and it seems a fair inference that the large eddies are of this general character, *i. e.* are of simple structure and are distributed along the direction of flow

Fig. 4.

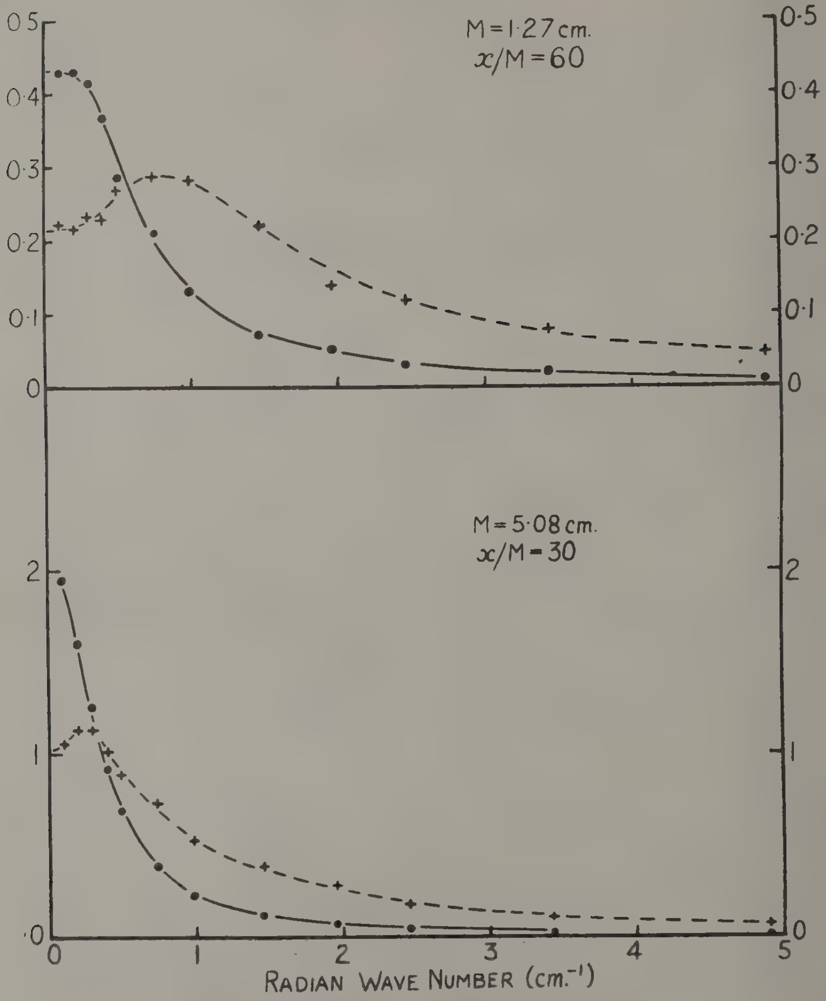


Spectrum functions of $\frac{u \pm v}{\sqrt{(2)}}$ components.

$$\left[\begin{array}{l} \bullet\ \frac{u+v}{\sqrt{(2)}} \\ +\ \frac{u-v}{\sqrt{(2)}} \end{array}\right]$$

randomly and not in a periodic or quasi-periodic manner. The initial surmise that the large eddies might be similar to a vortex-street must be discarded, and it is necessary to study the eddies more closely.

Fig. 5.



Spectrum Functions in Isotropic Turbulence.

[● *u*-component
+ *v*-component]

Further information can be obtained from the spectrum of the *v*-component of the velocity fluctuation. An assembly of randomly oriented simple eddies has a spectrum function for *v* which has a maximum at a non-zero frequency, and this is confirmed both by the spectrum of the v-component in isotropic turbulence, and of the large wave-number components of wake spectra. However, the low wave-number components of the wake spectra are again similar in form to a Gaussian error function, and it is concluded that the large eddies must have a preferred orientation. This preferred orientation of the axis of vorticity cannot be in the yOz plane, or the spectrum maxima would be very pronounced, and, since the u-component of the large eddies is not small, the preferrred orientation is not along Ox. This suggests that one the of principal axes of rate-of-strain of the mean flow (which are at ±45° to the Ox and Oy axes) may be the preferred direction, and this hypothesis is confirmed by measuring the spectra of the velocity components in these directions and along Oz. Considerable intensity of the low wave-number component is found in the spectra of w, and (in the positive half of the wake) of the $(u-v)/\sqrt{2}$ velocity component, while the intensity is much less for $(u+v)/\sqrt{2}$, *i. e.* along the axis of positive rate-of-strain. The preferred direction for the vorticity of the large eddies is approximately the direction of maximum positive rate-of-strain, which suggests that the source of the energy of the large eddies is the stretching they undergo as they are convected by the mean stream.

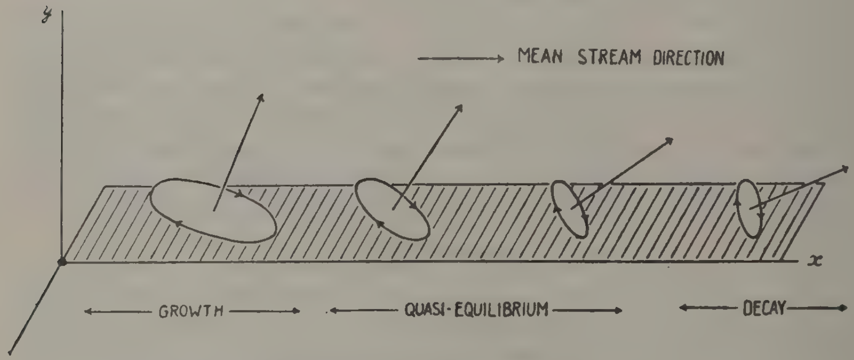
This examination of the wake spectra shows that the large eddies are of simple structure, are randomly placed in the direction of the free stream, and have a preferred orientation roughly along the principal axis of positive rate-of-strain. By transforming the spectra into the longitudinal correlation function, it can also be shown that the average extent in the Ox direction of a large eddy is considerably larger than the width of the wake, and the typical large eddy must be rather elongated in this direction.

§8. MECHANISM OF THE SLOW EDDIES.

Accepting the structure of the large slow eddies that has been outlined above, it is easy to see that such eddies receive energy from the mean flow, and may use it partly in increasing their kinetic energy and partly in working against the shear stresses caused by the smaller eddies of the main body of turbulence. This gain of energy cannot persist indefinitely for, after a sufficient time, the stretching action will cease and the large eddy will be dissipated (fig. 6). The typical history of a large eddy may be something like this. Initially, appreciable circulation occurs by chance in a circuit with normal inclined at a small angle to the Oy axis, and, as the circuit is convected by the mean flow, the energy increases as the area of the circuit decreases. In the early stages of development, it is likely that the dimensions of such an eddy are too great for the effect of eddy viscosity to be appreciable, but, as the circuit area decreases, the velocity gradients tend to increase and loss by eddy viscosity increases in importance. At the intermediate stage of development, it is probable that the energy gain from the mean flow and the energy loss by eddy viscosity are comparable in magnitude, and that the energy is nearly stationary. When the circuit area can no longer decrease, the eddy loses energy and disappears.

The viscosity control mechanism which has been described previously can easily be modified to fit into this scheme. When an eddy reaches an appreciable intensity, its convective action will lead to local spreading of the wake and to increased transfer of energy from the mean flow to the small-scale turbulent motion. The resulting increase in turbulent intensity and eddy viscosity reduces and finally stops the growth of the large eddy, and it may be supposed that, during the intermediate period of the eddy development, it is near a state of equilibrium such that the rate-of-change of total energy is small compared with the gain from the mean flow or the loss by eddy viscosity. The aggregate effect of a succession of these eddies is that the eddy viscosity is maintained continuously at a value such that the mean eddy is in this condition of dynamic equilibrium, the stabilizing process being: increase of energy of large eddy ⟶ increased rate of spread of wake ⟶ increased rate of production of small-scale turbulent energy ⟶ increasing eddy viscosity ⟶ increased energy loss from the large eddy. The time-scales for all these processes are comparable, and it seems very probable that a control process of this kind does operate in the wake.

Fig. 6.



Development of large eddy.

Confirmation of this notion can be obtained by inquiring whether eddies similar to the large eddies of the wake could be in a condition of dynamic equilibrium. Experimentally, the large eddies are found to be long cylindrical structures, generally similar to the eddy described by

$$u=v=Az\exp\{-\tfrac{1}{2}\alpha^2(y^2+z^2)\},$$
$$w=-Ay\exp\{-\tfrac{1}{2}\alpha^2(y^2+z^2)\},$$

which is an eddy with vorticity roughly in the direction $(1/\sqrt{2}, -1/\sqrt{2}, 0)$ and unlimited in the Ox direction. In the absence of better information about the exact structure of a typical large eddy, let us consider the energy balance of such an eddy in a mean velocity field of constant shear $\partial U/\partial y$. Assuming the eddy viscosity ϵ to be isotropic, the loss of energy by eddy viscosity is

$$D=\iint_{-\infty}^{\infty} W\,dy\,dz,$$

where

$$W=\rho\epsilon\left[2\left(\frac{\partial u}{\partial x}\right)^2+2\left(\frac{\partial v}{\partial y}\right)^2+2\left(\frac{\partial w}{\partial z}\right)^2+\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}\right)^2\right.$$
$$\left.+\left(\frac{\partial w}{\partial z}\right)+\left(\frac{\partial u}{\partial z}\right)^2\right],$$

and then

$$D=3\pi\rho\epsilon A^2\alpha^2.$$

The rate of production of energy is

$$S=\iint_{-\infty}^{\infty}\rho uv\frac{\partial U}{\partial y}\,dy\,dz$$
$$=\frac{\pi}{2}\rho A^2\frac{\partial U}{\partial y},$$

If the energy is stationary, $S=D$, and the equilibrium equation is

$$\frac{\partial U}{\partial y}=6\epsilon\alpha^2.$$

However, it may be argued that in a developing wake the equilibrium condition should be that the ratio of the eddy energy to the wake energy remains constant in the period of dynamic equilibrium. On this hypothesis, the equilibrium equation becomes

$$\frac{\partial U}{\partial y}+\frac{3}{4}\left[\frac{U_0}{x}\right]=6\epsilon\alpha^2,$$

where U_0/x is the time taken for the mean stream to travel from the virtual origin of the wake to the point of observation. From observations on the wake at a cylinder Reynolds number of 1430,

$$\left(\frac{\partial U}{\partial y}\right)_{\max}=3{\cdot}0\left[\frac{U_0}{x}\right]$$

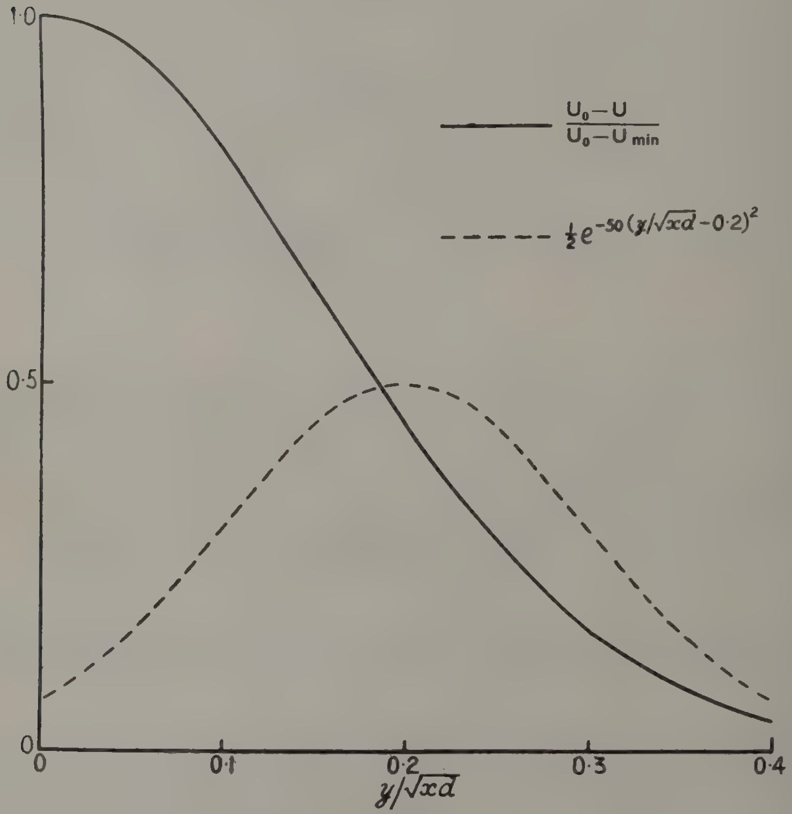
so the two forms of the equilibrium equation are not very different. At the same Reynolds number, it has been found that the mean value of the eddy viscosity is approximately 1/80 U_0d (Townsend 1950). Under these conditions the equilibrium size of a large eddy is given by

$$\alpha^2xd=40 \text{ or } 50,$$

depending on which of the two definitions of dynamic equilibrium is used. In fig. 7, the function exp. $(-\alpha^2y^2)$, representing roughly the distribution of energy in the large eddy, and the mean velocity defect U_0-U are plotted on the same scale, and the equilibrium large eddy is seen to fit very well into one half of the wake.

This very crude calculation shows that eddies of the dimensions necessary to explain the slow mixing processes would be approximately in a state of dynamic equilibrium, and that the suggested viscosity control process is possible. In free turbulence, the occurrence of intermittently turbulent flow[*] makes the existence of the controlling system of eddies especially obvious, but there is quite a possibility that similar processes may be at work in closed flows. The only experimental evidence for a double structure of turbulence in a closed flow is the observation of MacPhail (1944) that, in the fully turbulent flow between rotating cylinders, there is superimposed on the small-scale turbulent motion a quasi-periodic velocity variation of scale similar to the original laminar instability described by G. I. Taylor (1923). This velocity variation probably controls the level of eddy viscosity in a similar way to the large eddies of the wake flow. The contrast between the periodicity of the large eddies in cylinder flow and the aperiodic nature of the wake eddies is probably due to the constancy in

Fig. 7.



Wake profile and eddy size.

time of the cylinder flow which allows the development of a periodic instability. In wake flow, before a periodic instability has developed appreciably, the wake dimensions have increased by too much to permit the continued growth of the original instability.

§ 9. THE CONCEPT OF EDDY VISCOSITY.

This description of the turbulent motion in the cylinder wake in terms of an eddy viscosity has the advantage that, assuming the dynamic equilibrium of the large eddies, the exact mechanism of the production of eddy viscosity is irrelevant, and, in principle, it is possible to estimate the equilibrium eddy viscosity without considering the energy balance of the turbulent motion. There is no strong experimental evidence against this use of a coefficient of eddy viscosity to describe the small scale friction, for the scale and intensity of the small-scale turbulent motion are such that "time of flight" and "mixing length" are small compared with times and lengths associated with the mean flow. What is certainly incorrect is the attempt to relate its value to the local parameters of the mean flow, implying local and instantaneous energy equilibrium. The evidence shows that the setting up of Reynolds stresses must be a fairly rapid process compared with the attainment of turbulent energy equilibrium, *i. e.* that transfer of energy from one velocity component to another under the action of the shearing must be rapid. The comparative slowness of energy transfer from the mean flow to the turbulence and then to heat suggests that viscosity is irrelevant to the process of stress production, and, for similar reasons, turbulent transfer between eddies of different sizes also must have little effect on the establishment of the shear stress. The problem then reduces to the distortion of a turbulent field. If the axes of reference are parallel to the principal axes of strain, then the effect of an instantaneous distortion described by contradiction ratios l, m, n, is that

$$\xi_1 = l\xi_0, \quad \eta_1 = m\eta_0, \quad \zeta_1 = n\zeta_0 ,$$

where $lmn=1$ and ξ_1, η_1, ζ_1 are the final vorticity components ξ_0, η_0, ζ_0 are the initial vorticity components.

For the wake, $n=1$, and if ξ is in the direction of positive strain, $\overline{\xi^2}$ increases while $\overline{\eta^2}$ decreases. In terms of the corresponding velocity components, $\overline{u^2}$ decreases and $\overline{v^2}$ increases, leaving $\overline{w^2}$ unchanged, at least for small strains. This results in the production of a Reynolds shear stress, $\rho(\overline{u^2}-\overline{v^2})$, set up instantaneously and proportional to the total strain, and this is probably the first stage in the actual process. When the Reynolds stress developed by straining becomes large, the tendency of ordinary turbulence to isotropy should resist the tendency of the strain, to align the vorticity along the direction of position strain, and the eddy viscosity may represent the balance between the orientating action of the strain and the disorientating effect of the turbulent transfer between the velocity components. That the selective action of the strain acts on the vorticity rather than on the velocity makes it clear why the turbulent motion in free turbulence is apparently " isotropic." The intensities of the velocity components along the conventional axes of reference are nearly equal, because these axes are at 45° to the principal axes of strain, and transfer between the vorticity components along the principal axes hardly affects the energy in the velocity components along the ordinary axes of reference.

This notion that the eddy viscosity is a parameter describing a balance set up between the orientating effect of the shear and the disorientating effect of turbulent transfer is the last remaining possibility within the "equilibrium" group of theories of the eddy viscosity (after energy equilbrium has been discarded), and its accuracy is dependent on a sufficiently rapid attainment of this equilibrium. Revelant experiments would be on the attainment of isotropy in a uniform field of, say, axisymmetric turbulence, but existing measurements are not sufficiently comprehensive and it is proposed to investigate this matter directly. If the tendency to isotropy is too sluggish to be an effective balance to the straining process, it may be necessary to discard the equilibrium concept of shear stress production altogether, and to suppose that, like the energy, the shear stress is determined mostly by the previous history, probably by the total strain. While it would be premature to consider this hypothesis in detail, it should be mentioned that most of the conclusions concerning the mechanism of control of the large eddies would be equally valid, although the concept of eddy viscosity would then have less physical significance.

REFERENCES.

GOLDSTEIN, S, 1938, *Modern Developments in Fluid Dynamics* (Oxford : Clarendon Press)

MACPHAIL, D. C., 1944, *Royal Aircraft Establishment Report* No. 1928

TAYLOR, G. I., 1923, *Phil. Trans.* A, **223**, 289-343

TOWNSEND, A. A., 1948, *Australian J, Sci. Res.* A, **1**, 161-174 ; 1949, *Proc. Roy, Soc.* A, **197**, 124-140 ; 1950, *Australian J. Sci. Res.* (in the press).

LXXXII. *The Effect of Eddy Viscosity on Ocean Waves.*

By K. F. Bowden,
Oceanography Department, University of Liverpool *.

[Received May 22, 1950.]

Abstract.

The effect of turbulence on ocean waves appears to call for an eddy viscosity which, while being large compared with the molecular viscosity, is small compared with the eddy viscosity found to apply to currents. A recent suggestion that the eddy viscosity should be taken as proportional to $\lambda^{4/3}$ where λ is the wavelength, seems to the author to be unsatisfactory. It is proposed, instead, that the coefficient of eddy viscosity N, applicable to waves, should be of the form $N=Kca$, where c is the speed of propagation, a is the amplitude and K is a constant. This form is indicated on dimensional grounds and is shown to be in conformity with v. Karman's similarity hypothesis for shearing flow. With K of the order of 5×10^{-5}, the eddy viscosity would be large enough to account entirely for the observed rate of decay of ocean swell, while its effect on the attenuation of waves with depth would still be negligible. In the initial formation of waves, only the molecular viscosity need be considered. A possible effect of eddy viscosity on waves in the generating area would be to limit the steepness of the longer waves to a value less than the breaking steepness, even when the duration and fetch were unlimited.

§ 1. Introduction.

While molecular viscosity appears to be an important factor in the initial formation of surface waves on water, its dissipative effect on the long waves of ocean swell, for example, is negligibly small. If an eddy viscosity of the order of magnitude found to apply to ocean currents, on the other hand, were operative in the case of waves, their loss of energy would be much more rapid than is observed. Since it is well recognized in meteorology and oceanography that the eddy viscosity applicable to a particular motion depends on its scale, the suggestion has been made that waves may be affected by turbulence but that a smaller, and probably variable, eddy viscosity is applicable.

§ 2. Turbulence and Eddy Viscosity.

The first quantitative expression of this idea would appear to be that made recently by Groen and Dorrestein (1950), who assumed an eddy viscosity proportional to $\lambda^{4/3}$, where λ is the wavelength. This form was based on a result found by v. Weizsäcker (1948) in the theory of locally

---

* Communicated by the Author.

isotropic turbulence, and on an empirical law found by Richardson (1926) from diffusion experiments. V. Weizsäcker was considering the transfer of energy from turbulence on a scale characterized by a linear dimension l to turbulence of scales smaller than l. By similarity considerations, he found the effective eddy viscosity to be proportional to $1^{4/3}$ for values of l within a certain range. Turbulence on a scale so small that dissipation by molecular viscosity is the dominating factor, or so large that transfer of energy from the mean motion to the turbulence predominates is excluded from the range to which the similarity law applies. In the case of the dissipation of waves, we are dealing with the loss of energy from the mean motion (*i. e.* the wave motion) and it is not clear, therefore, that the eddy viscosity law deduced by v. Weizsäcker's theory will apply to this problem.

Richardson's result, which was obtained originally for the atmosphere and has been found by Richardson and Stommel (1948) to apply also to the ocean in certain conditions, was that the coefficient of diffusion applicable to the dispersal of a group of particles is proportional to $d^{4/3}$, where d is the separation of a pair of particles. The process considered in this case can take place in turbulence which does not necessarily involve shear in the fluid and it is not apparent that the dissipation of energy in a problem involving shearing stresses can be deduced from it.

A further difficulty is that, even if we assume an eddy viscosity proportional to $l^{4/3}$, the question arises : what is the characteristic length l for wave motion : Groen and Dorrestein have taken it to be λ, but one might expect the amplitude a also to be a characteristic length for a given set of waves.

§ 3. Wave Motion as affected by Molecular and Eddy Viscosity.

Let us consider two-dimensional surface waves in deep water. Let rectangular axes be taken, with Ox and Oy in the undisturbed sea surface and Oz vertically upwards. Let the waves be travelling in the direction of the negative x-axis, and U and W be the horizontal and vertical components of the particle velocity at any point. Then for irrotational waves of small amplitude

$$\left.\begin{aligned} \mathrm{U}&=-\omega a e^{kz}\cos(\omega t+kx),\\ \mathrm{W}&=-\omega a e^{kz}\sin(\omega t+kx),\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\omega=2\pi/\mathrm{T}$, T being the period, $k=2\pi/\lambda$, λ being the wavelength and a is the amplitude of the surface elevation. t is the time.

The rate of dissipation of energy by molecular viscosity may be shown to be (Lamb 1932, p. 623)

$$\mathrm{D}=2\mu k^3c^2a^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where μ is the coefficient of viscosity and $c=\lambda/\mathrm{T}$ is the velocity of propagation. If no external forces are acting, the amplitude of the waves will decay at a rate given by the factor exp $(-2\nu kt)$, where $\nu=\mu/\rho$ is the kinematic viscosity, ρ being the density.

To a first approximation, the components of the particle velocity are still given by (1), provided that

$$\theta \equiv \frac{\nu k}{c} \ll 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

By inserting numerical values in (3), we find that θ is, in fact, negligibly small, except for very short waves.

To consider the effect of eddy viscosity, we note that a wave motion is specified by the three quantities λ, T and a. We will assume, therefore, that if there exists an eddy viscosity N, applicable to wave motion, it is a function of λ, T and a. If we assume a power law, $N = K\lambda^{\alpha} a^{\beta} T^{\gamma}$ where K is a non-dimensional constant, then, since the dimensions of N are L^2T^{-1}, we find

$$\alpha + \beta = 2, \qquad \gamma = -1.$$

Taking $\alpha = \beta = 1$ as the simplest solution including λ and a,

$$N = K\frac{\lambda a}{T} = Kca. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

It is suggested that equation (4) gives the appropriate form of eddy viscosity to apply to wave motion. We need not expect K to be an absolute constant but rather a non-dimensional parameter which varies only slowly compared with the other quantities involved. The above treatment is still useful if we can take K as constant over a limited range of the quantities λ, a and T.

The rate of dissipation of energy by eddy viscosity, replacing μ by ρN in equation (2), becomes,

$$W = 2\rho K k^3 c^3 a^3. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

The particle velocity components will still be given to a first approximation by (1), provided that the criterion (3) is satisfied with N replacing ν, *i. e.*

$$\theta_N \equiv \frac{Nk}{c} = Kka \ll 1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (6)$$

It will be shown later that, with a value of K large enough for eddy viscosity to account entirely for the observed rate of decay of ocean swell, the condition (6) is still satisfied.

§ 4. V. KARMAN'S SIMILARITY HYPOTHESIS APPLIED TO WAVE MOTION.

An alternative approach to the problem of the dissipation of wave motion by turbulence may be made by v. Karman's similarity hypothesis for shearing flow, which has found successful application to turbulent flow through pipes and channels. Let u, v and w be the components of turbulent velocity and let square brackets [ ] denote values of functions of these components. Then if the mean velocity U is a function of z only, while $V = W = 0$, the rate of dissipation of energy of the mean motion by turbulence is given by

$$\psi = \rho[uw]\frac{dU}{dz}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (7)$$

v. Karman assumed dynamical similarity in the turbulence at various points in the field of motion and deduced that

$$[uw]=-l^2\left|\frac{d\mathrm{U}}{dz}\right|\frac{d\mathrm{U}}{dz} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where

$$l=\mathrm{K}_1\left|\frac{d\mathrm{U}}{dz}\right|\bigg/\left|\frac{d^2\mathrm{U}}{dz^2}\right| \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where K_1 is a non-dimensional constant.

In the case of two-dimensional motion, in the xz-plane, (7) is replaced by (Lamb, *loc. cit.*, p. 676)

$$\psi=\rho\left\{[u^2]\frac{\partial\mathrm{U}}{\partial x}+[w^2]\frac{\partial\mathrm{W}}{\partial z}+[uw]\left(\frac{\partial\mathrm{U}}{\partial z}+\frac{\partial\mathrm{W}}{\partial x}\right)\right\}. \quad . \quad . \quad (10)$$

In applying this equation to wave motion, we assume that mean values of $\partial\mathrm{U}/\partial x$ etc., as well as $[u^2]$ etc. are taken over an element of volume of dimensions small compared with the wavelength and over a time small compared with the period of the wave motion. If U and W are given by equations (1),

$$\psi=\rho k\omega a e^{kz}\{([u^2]-[w^2])\sin(\omega t+kx)-2[uw]\cos(\omega t+kx)\}. \quad . \quad . \quad (11)$$

By analogy with (8), we should expect $[uw]$ to be a function of $\partial\mathrm{W}/\partial x$ as well as of $\partial\mathrm{U}/\partial z$. In this case, however, $\partial\mathrm{W}/\partial x=\partial\mathrm{U}/\partial z$ so that we may take

$$[uw]=-l_4^2\left|\frac{\partial\mathrm{U}}{\partial z}\right|\frac{\partial\mathrm{U}}{\partial z},$$

where

$$l_4=\mathrm{K}_4\left|\frac{\partial\mathrm{U}}{\partial z}\right|\bigg/\left|\frac{\partial^2\mathrm{U}}{dz^2}\right|=\frac{\mathrm{K}_4}{k},$$

where K_4 is a constant.

Similarly, we will take

$$[u^2]=\mathrm{C}_2+\frac{\mathrm{K}_2^2}{k^2}\left|\frac{\partial\mathrm{U}}{\partial x}\right|\frac{\partial\mathrm{U}}{\partial x}$$

and

$$[w^2]=\mathrm{C}_3+\frac{\mathrm{K}_3^2}{k^2}\left|\frac{\partial\mathrm{W}}{\partial z}\right|\frac{\partial\mathrm{W}}{\partial z},$$

where K_2 and K_3 are constants, introduced in a way similar to K_4. C_2 and C_3 are positive constants, since $[u^2]$ and $[w^2]$ are essentially positive. From (11), we have, for the mean rate of loss of energy per volume from the wave motion at a given depth z,

$$\begin{aligned} d\mathrm{W} &= \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\psi\,dt \\ &= \rho k\omega^3a^3e^{3kz}\{(\mathrm{K}_2^2-\mathrm{K}_3^2)\overline{|\sin(\omega t+kx)|\sin^2(\omega t+kx)} \\ &\qquad +2\mathrm{K}_4^2\overline{|\cos(\omega t+kx)|\cos^2(\omega t+kx)}\} \\ &= \frac{4}{3\pi}\rho k\omega^3a^3e^{3kz}(\mathrm{K}_2^2-\mathrm{K}_3^2+2\mathrm{K}_4^2). \end{aligned}$$

For the rate of dissipation per unit area of the sea surface,

$$W=\int_{-\infty}^{0} dW\,dz=\frac{4}{9\pi}\,\rho\omega^3a^3(K_2^2-K_3^2+2K_4^2). \quad . \quad . \quad . \quad (12)$$

Hence

$$W=2\rho K k^3c^3a^3$$

as in equation (5) if

$$K=\frac{2}{9\pi}\,(K_2^2-K_3^2+2K_4^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The equations of wave motion in the xz-plane, in the presence of turbulence are (from Lamb, *loc. cit.*, p. 678)

$$\frac{\partial U}{\partial t}+\frac{1}{\rho}\frac{\partial p}{\partial x}+\frac{\partial}{\partial x}\,[u^2]+\frac{\partial}{\partial z}\,[uw]=0\,,$$

$$\frac{\partial W}{\partial t}+\frac{1}{\rho}\frac{\partial p}{\partial z}+\frac{\partial}{\partial z}\,[w^2]+\frac{\partial}{\partial x}\,[uw]=g\,,$$

where p is the wave pressure. To the first approximation

$$p=\rho gae^{kz}\cos(\omega t+kx)\,.$$

The condition that equation (1) should still be valid, in the presence of turbulence, to a first approximation, is that terms such as $\partial[uw]/\partial z$ should be small compared with terms such as $(1/\rho)\partial p/\partial x$.

We have $\left|\frac{\partial}{\partial z}\,[uw]\right|=K_4^2k\omega^2a^2e^{2kz}$ and $\left|\frac{1}{\rho}\,\frac{\partial p}{\partial x}\right|=gkae^{kz}$. Since K_4^2 is of the same order of magnitude as K, we may write the condition as

$$\frac{Kk\omega^2a^2e^{2kz}}{gkae^{kz}}\ll 1$$

and since $kc^2=g$, this becomes

$$Kkae^{kz}\ll 1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Apart from the factor $\exp kz$, condition (14) is the same as that given by (6).

We may therefore derive the equation for the dissipation of wave motion by turbulence from v. Karman's similarity hypothesis without introducing the concept of eddy viscosity or the analogy with molecular viscosity.

§ 5. The Decay of Ocean Swell.

5.1. *Waves of a Single Period.*

We will consider waves of constant wave velocity c, travelling in the direction Ox and assume a quasi-stationary state, *i. e.* the wave energy varying with distance but constant in time at a given point. Then if E is the energy per area of the waves, the rate of loss of energy with distance, due to eddy viscosity, would be given by

$$\frac{c}{2}\,\frac{dE}{dx}=-W \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

assuming the energy to be propagated with the group velocity $c/2$. Since $E=\frac{1}{2}\rho g a^2$ and taking W from equation (5), we have

$$\frac{da}{dx}=-\frac{4Kg^2a^2}{c^4}$$

using the relation $kc^2=g$. The solution is

$$\frac{1}{a}=\frac{1}{a_0}+\frac{4Kg^2}{c^4}x, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where a_0 is the amplitude at $x=0$. In terms of the period T, since $gT=2\pi c$,

$$\frac{1}{a}=\frac{1}{a_0}+\frac{64\pi^4K}{g^2T^4}x. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Sverdrup and Munk (1947), in their treatment of the growth and decay of waves, assumed that the effect of eddy viscosity was negligible and attributed the decay of swell to air resistance, the rate of dissipation per area being of the form

$$R=\tfrac{1}{2}s\rho' k^2c^3a^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where s is a coefficient and ρ' the density of the air. If we replace W by R in equation (15), we are led to a solution

$$a=a_0 \exp\{-(s\rho' g/\rho c^2)x\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

or in terms of T

$$a=a_0 \exp\{-(4\pi^2 s\rho'/\rho g T^2)x\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

A comparison of equations (17) and (20) shows the differences to be expected if the decay is due mainly to eddy viscosity or to air resistance. Both equations show that the decay is more rapid for waves of shorter period, the discrimination being greater in the case of eddy viscosity. In either case, therefore, if a spectrum of periods is present in the swell at $x=0$, the maximum amplitude occurring for a period T_0, then at a distance x, owing to the selective dissipation, the component of maximum amplitude will occur at a period T where $T>T_0$. An increase of effective period with distance usually occurs with ocean swell, but Munk (1947) has shown that such an increase may be explained, in some cases at least, on kinematical grounds without any selective dissipation of energy.

5.2. *Effective Period increasing with distance.*

Let us consider the general case of the period T, and thus the velocity c, increasing with x, without assigning the cause, and follow Sverdrup and Munk by assuming our dissipation equation to apply to the " significant waves ". If c varies with x, equation (15) is replaced by

$$\frac{1}{2}\left(c\frac{dE}{dx}+E\frac{dc}{dx}\right)=-W. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Hence

$$\frac{da}{dx}=-\frac{1}{2}\frac{a}{c}\frac{dc}{dx}-\frac{4Kg^2a^2}{c^4}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Little is known of the way in which c may be expected to increase with x, but to obtain a solution we will assume that

$$\frac{dc}{dx}=\sigma \text{ (a constant), so that } c=c_0+\sigma x.$$

Then the solution of (22) may be shown to be

$$\frac{1}{ac^{1/2}}=\frac{1}{a_0c_0^{1/2}}+\frac{8\mathrm{K}g^2x}{7(c-c_0)}\left(\frac{1}{c_0^{7/2}}-\frac{1}{c^{7/2}}\right), \quad . \quad . \quad . \quad . \quad . \quad (23)$$

where a_0, c_0 are values at $x=0$ and a, c are values at x. Putting $g\mathrm{T}=2\pi c$,

$$\frac{1}{a\mathrm{T}^{1/2}}=\frac{1}{a_0\mathrm{T}_0^{1/2}}+\frac{128\pi^4\mathrm{K}x}{7g^2(\mathrm{T}-\mathrm{T}_0)}\left(\frac{1}{\mathrm{T}_0^{7/2}}-\frac{1}{\mathrm{T}^{7/2}}\right). \quad . \quad . \quad . \quad (24)$$

For dissipation by air resistance, it may be shown that the corresponding solution is

$$a\mathrm{T}^{1/2}=a_0\mathrm{T}_0^{1/2}\exp\{-(4\pi^2s\rho'/\rho g\mathrm{T}_0\mathrm{T})x\}. \quad . \quad . \quad . \quad . \quad (25)$$

5.3. *Application to Observations of Swell.*

The above equations have been applied to the examples of swell given by Sverdrup and Munk (*loc. cit.*, Table IV.). s and K have been computed from the recorded values of x, a_0, T_0, a and T, and are shown in Table I. For the examples of swell Nos. 1, 2, 3 and 5, equations (24) and (25) respectively were used. For Nos. 4 and 6, equations (17) and (20) were used, since T differs little from T_0. The scatter of the values of s and K is about the same and one cannot decide on these data whether the decay law for air resistance or eddy viscosity fits the observations better. It is possible, of course, that both processes operate and are of the same order of magnitude.

Taking the mean value of K, we would have

$$\mathrm{N}=5{\cdot}6\times10^{-5}ca. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

As an example, for swell of wave velocity 10 m./sec. and amplitude 2 m., $\mathrm{N}=11{\cdot}2$ cm.^{2}/sec., which does not appear unreasonable. Referring to the condition given by (6), we see that, with this value of K, the condition is easily satisfied. The influence of eddy viscosity on the rate of attenuation of the waves with depth is, therefore, negligible.

§ 6. THE INITIAL FORMATION OF WAVES.

Jeffreys (1925) showed that the early stages of the formation of waves could be explained if a transfer of energy from wind to waves took place through normal pressures on the wave profile. The rate of transfer of energy was taken to be of the form

$$\mathrm{A}=\tfrac{1}{2}s\rho'(ka)^2(\mathrm{V}-c)^2c \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where $\mathrm{V}>c$, s being the “ sheltering coefficient ” and V the wind speed. Comparing A with D, the rate of dissipation by molecular viscosity, Jeffreys showed the existence of a critical value of V, below which waves could not be formed. From observations he found $V_{\min}=110$ cm./sec. (approximately), which corresponds to $s=0\cdot27$. If $V>V_{\min}$ this treatment sets no limit to the height the waves may attain and they would grow, presumably, until they reached the breaking steepness of $2a/\lambda=1/7$.

From wind tunnel measurements on wave models of various ratios of $2a/\lambda$, Motzfeld (1937) deduced that A was of the form

$$A=\tfrac{1}{2}s'\rho'(ka)^{3/2}(V-c)^2c \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

TABLE I.

Decrease in Height of Swell.

1	2	3	4	5	6	7	8
No.	Decay distance x	At end of fetch		At distance x		Air resistance s	Eddy viscosity K
		Ampl. a_0	Period T_0	Ampl. a	Period T		
	km.	m.	sec.	m.	sec.	$\times10^{-2}$	$\times10^{-5}$
1	2,320	3·4	7·9	0·7	11·5	1·04	4·3
2	1,480	2·4	8·1	1·3	17·0	0·46	2·2
3	1,950	3·45	8·7	1·0	17·0	1·39	7·1
4	1,110	2·35	10·5	1·35	11·2	1·20	6·1
5	1,480	2·55	12·2	1·5	9·1	1·04	4·6
6	185	1·25	9·0	1·05	9·5	1·84	9·3
					Mean	1·16	5·6

The data in columns 1 to 6 are taken from Table IV. of Sverdrup and Munk's paper (1947). Column 7 gives the value of s, computed from equation (20) or (25), assuming the decay to be due to air resistance only. Column 8 gives the value of K, computed from equation (17) or (24), assuming the decay to be due to eddy viscosity only.

with $s'=0\cdot014$. Applying this result to the formation of waves, he found no critical wind speed in the region of 100 cm./sec. but for any value of V, ka is a definite function of c, which is a maximum when $c=V/3$. For $V<100$ cm./sec., the maximum ka would be negligibly small, but would increase rapidly in the region just above 100 cm./sec.

Neumann (1949), after considering observations of wind stress and wave steepness for various wind speeds, has proposed an equation for A of the form

$$A=\tfrac{1}{2}s''\rho' ka(V-c)^2c \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

with $s''=0{\cdot}095$. Like Motzfeld, he found $ka_{\max}$ to be a definite function of V and to increase rapidly when V was about 100 cm./sec., but his values of *ka* were appreciably higher than Motzfeld's.

While differing in detail, therefore, all investigators agree that easily observable waves should first appear when V=100 cm./sec. approximately, and have a velocity $c=V/3$, *i. e.* 33 cm./sec. approximately. If we take $a=0{\cdot}1$ cm., equation (26) would give $N=1{\cdot}85\times10^{-4}$ cm.^{2}/sec., *i. e.* much less than the kinematic molecular viscosity ν, which is $1{\cdot}8\times10^{-2}$ cm.^{2}/sec. We may conclude that eddy viscosity can play no part in the initial formation of waves and that only dissipation by molecular viscosity need be considered.

It seems unlikely that the relation $N=Kca$ would hold down to values of N comparable with ν. We might, in fact, define a " Reynolds' number for wave motion " by $R_w=ca/\nu$, and expect that turbulence would affect the wave motion only when R_w was above a certain value.

§ 7. The Growth of Waves under the Influence of Wind.

Although Sverdrup and Munk have used an expression for A similar to Jeffreys' (equation (27) for the transfer of energy from wind to waves by normal stresses for V>6 m./sec., they found it necessary to take a much lower value of *s*, i. e. $s=0{\cdot}013$. They showed that transfer of energy may also take place by tangential stresses, according to the equation

$$A'=\gamma^2\rho'(ka)^2V^2c, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where γ^2 is the resistance coefficient and was taken as $2{\cdot}6\times10^{-3}$. Unlike A, A′ remains finite when $c=V$. Dissipation by molecular viscosity is negligible for waves for which *c* is several m./sec. If eddy viscosity is also negligible (as assumed by Sverdrup and Munk), the height attainable by waves under the influence of moderate and strong winds is limited only by the breaking steepness, provided the duration and fetch are both sufficiently great. In practice, it is observed that, while the shorter waves may be continually breaking, waves travelling with speeds approaching that of the wind seldom reach the breaking steepness, and usually do not exceed about 1/3 of that steepness. On Sverdrup and Munk's theory, the explanation given is that sufficiently large durations and fetches do not, in fact, occur to enable the long waves to reach a greater steepness.

If eddy viscosity is the limiting factor, we see that the wave steepness cannot exceed an equilibrium value given by

$$A+A'=W.$$

Taking A from equation (27) and A′ from equation (30), we find

$$ka_{\max}=\frac{\rho'[s(1-\beta)^2+2\gamma^2]}{4\rho K\beta^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where $\beta=c/V$, and has been termed the " wage age ". Thus if s and γ^2 are constant, or are functions of β only, the equilibrium steepness is a function of the wave age, β, only. This general result is still true if A is given by either of the alternative equations (28) and (29).

From equation (31) the equilibrium steepness given by $ka_{\max}$ will be less than the breaking steepness, for which $ka=\pi/7$, if

$$\frac{\beta^2 K}{s(1-\beta)^2+2\gamma^2}>\frac{7\rho'}{4\pi\rho} \quad i.e. \quad >6\cdot8\times10^{-4}$$

inserting the numerical values of ρ' and ρ. We will consider two special cases of this condition :

(*a*) If $\beta=\frac{1}{3}$ and γ^2 is negligible,

$$K>2\cdot72\times10^{-3}s.$$

With Jeffreys' value, $s=0\cdot27$, $K>7\cdot3\times10^{-4}$. With Sverdrup and Munk's value, $s=0\cdot013$, $K>3\cdot5\times10^{-5}$, but this is hardly valid, since the value $s=0\cdot013$ was derived empirically on the assumption that eddy viscosity was negligible.

(*b*) If $\beta=1$, then

$$K>1\cdot36\times10^{-3}\gamma^2.$$

Taking $\gamma^2=2\cdot6\times10^{-3}$, $K>3\cdot5\times10^{-6}$.

Comparing these inequalities with the value of K deduced in § 5.3, it seems possible that eddy viscosity may limit the steepness of waves travelling with a speed approaching that of the wind, but unlikely that it will limit the steepness of waves travelling at about one-third of that speed.

A relation between wave steepness and wave age was, in fact, found empirically by Sverdrup and Munk. Until we have more data on the rate of transfer of energy from wind to waves, however, we cannot say whether the steepness of the longer waves is limited only by the duration and fetch, or by an eddy viscosity of the form we have been considering.

§ 8 THE EFFECT ON WAVES OF TURBULENCE DUE TO OTHER CAUSES.

It has sometimes been assumed that the turbulence affecting waves should be attributed to the effects of wind driven or other currents. Alternatively, it has been suggested that the turbulence affecting currents is derived from wave motion. Neither assumption is made in the present treatment, the turbulence being regarded as inherent in the wave motion itself. The wave motion generates the turbulence and the turbulence reacts on the wave motion. The question still arises, however, whether, in the presence of a shearing current, the increased turbulence would not greatly increase the effective eddy viscosity for waves. A clue to the answer may be found in equation (10), from which it is seen that for dissipation of energy by turbulence to occur, the mean values of terms such as $[uw]\partial U/\partial z$ must be finite. This implies not only a correlation between u and w, but also a correlation between $[uw]$ and $\partial U/\partial z$. It is assumed that, in the case of turbulence associated with the waves themselves, such correlations exist.

If we now consider the turbulence associated with, for example, a wind-driven current, there will undoubtedly be some components in the spectrum of turbulence of scale comparable with, or smaller than, the wavelength of the waves. These components will contribute to $[u^2]$ and $[w^2]$ in equation (10) and may possibly have a finite value of $[uw]$. It is only, however, if the contributions to $[u^2]-[w^2]$ and $[uw]$ have components which vary in phase with $\partial U/\partial x$ and $\partial U/\partial z$ respectively, that they will add anything to the dissipation of the waves. There is, therefore, no real discrepancy between the simultaneous existence of a large degree of turbulence due to other causes and a comparatively small eddy viscosity applicable to wave motion.

ACKNOWLEDGMENT.

It is a pleasure to record my thanks to Professor J. Proudman for his helpful criticism and suggestions in preparing this paper.

REFERENCES.

GROEN, P., and DORRESTEIN, R., 1950, *Nature, Lond.*, **165**, 445.
JEFFREYS, H., 1925, *Proc. Roy. Soc.* A, **107**, 189.
v. KARMAN, TH., 1930, *Proc. 3rd Internat. Congress for Applied Mechanics, Stockholm*, **1**, 85.
LAMB, H., 1932, *Hydrodynamics*, 6th edn. (Cambridge: University Press).
MOTZFELD, H., 1937, *Zeits. f. angew. Math. u. Mech.*, **17**, 193.
MUNK, W. H., 1947, *Trans. Amer. Geophys. Un.*, **28**, 198.
NEUMANN, G., 1949, *Deutsche Hydrog. Zeits.*, **2**, 187.
RICHARDSON, L. F., 1926, *Proc. Roy. Soc.* A, **110**, 709.
RICHARDSON, L. F., and STOMMEL, H., 1948, *J. of Meteorology*, **5**, 238.
SVERDRUP, H. U., and MUNK, W. H., 1947, U.S. Hydrographic Office, Pub. No. 601.
V. WEIZSÄCKER, C. F., 1948, *Zeits. f. Phys.*, **124**, 614.

LXXXIII. *Search for the Negative Proton in Fission,*

By J. M. C. Scott * and E. W. Titterton,
Atomic Energy Research Establishment, Harwell, Berks †.

[Received May 17, 1950]

During the 1946 Cambridge Conference a paper was read by Broda, Feather and Wilkinson (1947) describing ionization chamber and chemical experiments designed to test the possibility of emission of a negative proton as an alternative to the emission of a β-particle and a neutron, which occurs in certain fission products and is responsible for the delayed neutrons. Recently Kuan-Han Sun (1949), who appears unaware of the experiments of Broda *et al.*, has suggested that the delayed neutron emitter N^{17} might undergo negative proton decay. Still more recently, the situation has been reviewed by Feather (1949). The two competing reactions are:

$$X_Z^A \rightarrow Y_{Z+1}^{A-1} + (\beta^- + \nu) + n_0^1 + Q_1 \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$X_Z^A \rightarrow Y_{Z+1}^{A-1} + P^- + Q_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If the negative proton has the same mass as the positive proton $(H_1^1 - m_e c^2)$, then from the known mass values (*cf.* Stephens 1949)

$$Q_2 = Q_1 + 1{\cdot}79 \text{ MeV.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Since Q_1 is positive for the delayed neutron emitted in fisson, it follows that reaction (2) is energetically possible and non-relativistic negative protons might be emitted.

In the course of an investigation in this laboratory, involving the examination of large numbers of tracks formed by the slow neutron fission of U^{235} in photographic plates, a search has been made for associated events which might furnish evidence of negative proton emission from a fragment. The delayed neutron emitters have lives ranging from about 0·05 to 56 seconds (Hughes *et al.*, 1948) and, as the fragments come to rest in the photographic emulsion in a time less than 10^{-11} sec., it might be expected that the negative proton would be emitted after the nucleus has stopped. If a slow negative proton is emitted the question arises as to its ultimate fate.

Among the possibilities are:

(i) annihilation with a proton in the emulsion with emission of two γ-rays;

---

* Now at the Cavendish Laboratory, Cambridge.

† Communicated by the Authors.

(ii) capture by a heavy nucleus such as Ag or Br and subsequent annihilation with a proton of the nucleus with a chance that one of the γ-rays interacts with nucleons of the same nucleus leading to meson production and break-up of the nucleus to produce a “ star ”;

(iii) capture by a nucleus and conversion of the rest masses of the positive and negative protons into excitation energy of the nucleus in some other way, leading to a high energy nuclear explosion.

(iv) radiative capture by a nucleus which, unlike the first three, does not involve the annihilation of the “ rest mass ” of the particle.

The idea that negative protons may exist is derived ultimately from analogy with electrons. The hole theory, with its sea of occupied negative energy states, was first introduced as a complement to the Dirac electron theory and the discovery of the positron then fitted in naturally. The application of the same ideas to neutrinos is also generally accepted. In the case of protons, although the Dirac equation does not give the right magnetic moment, it nevertheless appears to be the proper way of describing particles with spin $\frac{1}{2}$. If the true relativistic theory of protons does include a hole theory and negative protons, it seems certain that they must be capable of annihilation with positive protons with emission of two γ-rays, just like positrons.

An objection to this view is as follows. If negative protons can be produced according to equation (2) which, omitting those nucleons not transformed, can be written :

$$N+N\rightarrow P^{+}+P^{-}+\text{k.e.} \qquad (4)$$

then it might appear that two neutrons in any nucleus can annihilate by way of virtual creation of a negative proton as an intermediate state

$$N+N\rightarrow P^{+}+P^{-}\rightarrow 2\gamma, \qquad (5)$$

and it follows that most known nuclei ought to explode. Thus if negative protons exist and are produced by Feather's process (2), it would seem likely that they cannot annihilate with positive protons, and therefore that they do not bear the same relation to protons as positrons do to electrons. This argument would only rule out the creation by Feather's process of a negative proton analogous to a positron. It is still possible to imagine a negative particle which cannot turn its rest energy into photons in this way. If created by Feather's process, it would then be absorbed by process (iv), *e. g.*

$$Br^{79}+P^{-}\rightarrow Se^{80}+\text{ energy} \qquad (6)$$

and gamma rays would be emitted similar to those from neutron capture. However, it does seem reasonable to postulate the existence of the anti-protons demanded by the hole theory, and meet the above objection by supposing that the neutrons occurring in nature can only react with anti-neutrons, and to assume that these (like positrons) do not normally occur in nature.

If we assume, therefore, that a negative proton can undergo annihilation with a proton, we may calculate the probability of the process by regarding it as the dropping of an ordinary proton into a hole. Even if the Dirac equation is not strictly valid for protons (in view of the magnetic moment difficulty) it seems to likely give the correct order of magnitude. The mathematical argument is the same as that for electrons, and gives a cross-section for negative protons (of velocity $v \leqslant c$)

$$\sigma = \pi e^4 / M^2 c^3 v, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and a lifetime of the order of 10^{-3} seconds in a material containing free protons. This time is much longer than that required for the particle to come to rest in the emulsion, so if the particle escapes from the fission fragment nucleus at all it ought not to be annihilated in flight.

These arguments lead to the view that possibilities (i) and (iv) above would allow a negative proton to be observed in the emulsion as a track of appreciable length emanating from the end of a fission track. Possibilities (ii) and (iii) on the other hand would give rise to a track leaving the end of a fission fragment and terminating in a star of tracks, some of which would be at low energy.

In the experiments which have been carried out in this Establishment some $6 \cdot 5 \times 10^5$ fission tracks representing $1 \cdot 3 \times 10^6$ fission fragments have been examined. About half were in Ilford C_2 emulsions sensitive to protons of energy up to 40 MeV and half in D_1 emulsions which, under the conditions of the experiment, recorded protons up to a maximum energy of 2·5 MeV. No tracks longer than 3μ (proton energy $\sim$300 kV) were observed associated with the ends of fission tracks though many knocked-on protons (identified from the geometry of the events) were observed along the fragment tracks. One case of a 3-particle star having its origin at the end of a fission track was shown to be a chance juxtaposition of events. It was identified as being similar to many other stars found in the emulsions resulting from the decay of chains of α-emitters present as contaminant in the uranium solutions employed for impregnating the plates.

From this result it appears that negative proton emission, if it occurs at all, is an improbable process. Accepting 0·8 per cent (Hughes *et al.*, 1948) as the yield of delayed neutron emitters in the slow neutron fission of U^{235} the experiment allows us to set an upper limit of 1 in 5×10^3 to the probability of negative proton emission relative to delayed neutron emission from these fragments.

We thank the Director, A.E.R.E., for permission to publish this paper.

REFERENCES

BRODA, FEATHER and WILKINSON, 1947, *Report on an International Conference on Fundamental Particles held at the Cavendish Laboratory, Cambridge, in July* 1946. (London : Published by the Physical Society) p. 114.
FEATHER, 1949, *Edinburgh Conference, November*, 1949.
HUGHES, DABBS, CAHN and HALL, 1948, *Phys. Rev.*, **73**, 111.
KUAN-HAN SUN, 1949, *Phys. Rev.*, **76**, 1266.
STEPHENS, 1949, *Phys. Rev.*, **76**, 181.

LXXXIV. *The Emission of Short Range Alpha Particles from Light Elements under Proton Bombardment.*—III. *The Reactions* ^{6}Li($p\alpha$)^{3}He *and* ^{7}Li($p\gamma$)^{8}Be*.

By W. E. Burcham and Joan M. Freeman,
Cavendish Laboratory, Cambridge[†].

[Received June 14, 1950.]

Summary.

The emission of short range alpha particles associated with the 14·8 MeV. gamma radiation produced in the ^{7}Li($p\gamma$) reaction has been established. The energy distribution of these particles can be interpreted as due to a level in the ^{8}Be nucleus at 3030±100 keV. energy, with a width of 2000 keV.

Some observations on the energy release (3·97±0·03 MeV.), cross section and excitation function for the ^{6}Li($p\alpha$)^{3}He reaction are also presented.

§1. Introduction.

The disintegration of ^{7}Li by protons is of particular interest since a well known resonance level of the compound nucleus ^{8}Be can easily be observed. The properties of this level have been studied in detail by Devons and Hine (1949) by observations on the capture reaction

$$^7\text{Li}+{}^1\text{H} \rightarrow {}^8\text{Be}^* \rightarrow {}^8\text{Be}+\gamma_1. \qquad (1)$$

The energy of the gamma radiation emitted by the excited ^{8}Be* nucleus formed in this reaction was measured by Walker and McDaniel (1948) and components of 14·8 MeV. and 17·6 MeV. were found. The energy 17·6 MeV. agrees with that expected if the ^{8}Be nucleus is left in its (unstable) ground state after the emission of the radiation. The 14·8 MeV. component could be due to a transition to an excited state of energy 2·8 MeV. in the ^{8}Be nucleus, in accordance with the reaction

$$^7\text{Li}+{}^1\text{H} \rightarrow {}^8\text{Be}^* \rightarrow {}^8\text{Be}^{**}+\gamma_2. \qquad (2)$$

Evidence from other reactions for the existence of such a state is summarized by Wheeler (1941), and a ^{8}Be nucleus left in this state is known to break up with the emission of two alpha particles of about 1·4 MeV. energy[‡]. It follows that short range alpha particles should be associated with the 14·8 MeV. gamma radiation observed in the Li($p\gamma$) reaction, providing that this radiation results from a transition to the alpha-emitting level of ^{8}Be. In the present work, of which a preliminary account has already been given (Burcham and Freeman 1949 a), such alpha particles were found and their energy distribution was examined and compared with that predicted from the gamma ray spectrum.

[†] Communicated by O. R. Frisch.

[‡] (*a*) The energy of the alpha particle pair exceeds the energy of the excited state by 89±5 keV., which is the disintegration energy of the ground state of ^{8}Be according to the measurements of Tollestrup, Fowler and Lauritsen (1949).

(*b*) Reaction (2) and the subsequent alpha particle emission might be referred to as the ^{7}Li($p\gamma'$,α)^{4}He reaction, in accordance with the terminology adopted by Fowler, Lauritsen and Lauritsen (1948), but in this paper the abbreviation ^{7}Li($p\gamma$)^{8}Be* will be used.

In the course of this work a number of check experiments were made on the disintegration of ^{6}Li by protons, mainly in order to assess the contamination by the lighter isotope of targets of separated ^{7}Li. The results of these observations are also presented.

§2. Experimental Method.

(a) *General.*

The apparatus used in these experiments has been previously described (Burcham and Freeman 1949 b, 1950; subsequently referred to as I, II). In some of the work the analysing magnet was turned on its side so that observations could be made in a horizontal plane and the target could be brought nearer the high voltage accelerating tube; larger currents could so be obtained.

(b) *Targets.*

Thin lithium targets were prepared by evaporating the metal *in vacuo* on to copper discs. During transference from the evaporator to the target chamber these targets lost their metallic lustre and were presumably converted into lithium hydroxide or carbonate† (*cf.* Bonner and Evans 1948 b). Targets of the ^{7}Li and ^{6}Li isotopes (similarly oxidized by exposure to air) were also available. These were deposited on copper, silver or aluminium backings in the small electromagnetic separator at the Atomic Energy Research Establishment, Harwell, and had surface densities, calculated from the ion current, of between 10 and 100 μgm./cm.^{2} of lithium before exposure to the atmosphere. A fairly satisfactory thin target of ^{7}Li was also prepared by converting a ^{7}Li sample into sulphate and evaporating *in vacuo* on to a copper button.

Target thicknesses in keV. equivalent energy for protons and alpha-particles were calculated assuming a stopping power of 400 keV./mgm./cm.^{2} for lithium carbonate for protons of 400 keV. energy. This value is consistent with the results of Warshaw (1949), with the figures given by Bonner and Evans (1948), and also with direct measurements made by comparison of the shape of the ^{7}Li($p\gamma$) resonance curve with that obtained from a thick target of the same composition (*cf.* Tangen 1946). The targets used were of thickness equivalent to 10–90 keV. energy for protons.

---

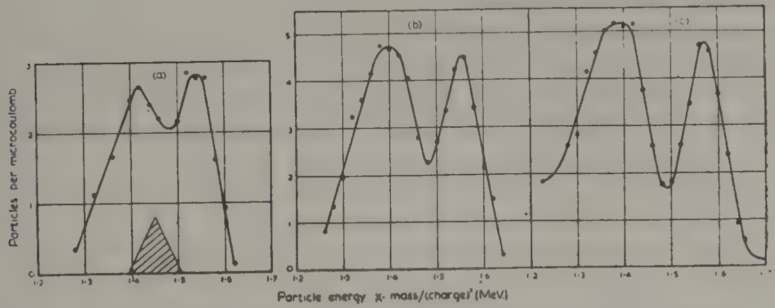
† The surface densities given refer to the oxidized deposit except where it is otherwise stated.

§3. RESULTS FOR ^{6}Li.

The disintegration of ^{6}Li by protons leads to the emission of short range alpha particles and ^{3}He nuclei according to the reaction

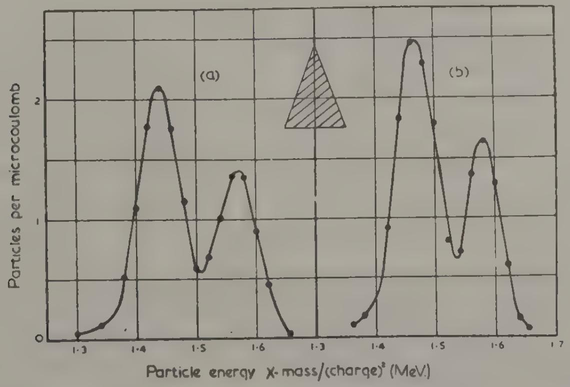
$$^6\text{Li}+{}^1\text{H} \rightarrow {}^3\text{He}+{}^4\text{He}. \quad \ldots\ldots\ldots \quad (3)$$

Fig. 1.



Alpha particles and ^{3}He nuclei from the reaction ^{6}Li($p\alpha$) ^{3}He. Angle of observation 135°. Proton energy (*a*) 330 keV., (*b*) 630 keV., (*c*) 930 keV. Target thickness 90μgm./cm.^{2}. The shaded triangle gives the analyser resolution.

Fig. 2.



Alpha particles and ^{3}He nuclei from the reaction ^{6}Li($p\alpha$)^{3}He. Angle of observation 135°. Proton energy (*a*) 380 keV., (*b*) 450 keV. Target thickness 15μgm./cm.^{2}. The shaded triangle gives the analyser resolution.

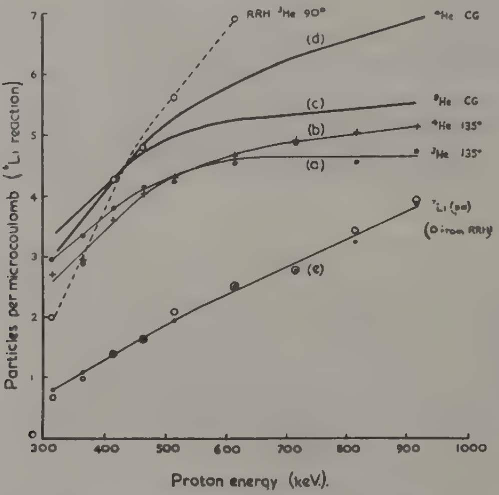
References to determinations of the energy release in this reaction and to observations on the cross section, excitation function and angular distribution are given by Hornyak and Lauritsen (1948) and a later determination of the enegy release as 4·017±·022 MeV. has been published by Tollestrup, Fowler and Lauritsen (1949).

Fig. 1 (*a*), (*b*), (*c*) shows the spectrum of particles obtained from an ordinary lithium target of surface density 90 μgm./cm.^{2}. The curves were all taken with an analyser resolving power of 50 (4 per cent resolution in energy) and at an angle of observation of 135° with the proton beam.

Bombarding energies of 930, 630 and 330 keV. were used, and the ^{6}Li($p\alpha$) yield was monitored by simultaneous observation of the alpha-particles from the ^{7}Li($p\alpha$)^{4}He reaction. Fig. 2 (*a*), (*b*) gives curves taken at the same angle, and with the same resolution, for bombarding energies of 380 and 450 keV. and with a target of surface density 15 μgm./cm.^{2}; these two curves were used for detailed comparison with the results obtained with thin ^{7}Li targets.

The two groups in figs. 1 and 2 are the ^{4}He and ^{3}He particles emitted in reaction (1). The energy of these particles observed at 135° does not vary greatly for the range of bombarding energies 300–900 keV. This simplified the determination of an excitation function for the reaction over this range, since it was only necessary to set the analyser field for the peak of either group and to take counts for a series of bombarding energies; checks were made at a number of points to ensure that the analyser field was correctly set. The yield of reaction (1) so observed for the 90 μgm./cm.^{2} target is shown in fig. 3; curves (*a*) (*b*) give the actual observations for the ^{3}He and ^{4}He particles, and curves (*c*) (*d*) are corresponding smoothed curves after correction for centre of mass motion. The angle in the mass-centre system for which the observations are actually made depends on the bombarding energy and varies from 139° to 143° over the energy range studied. The ^{3}He particles correspond to alpha particles emitted in the forward direction between angles of 30°

Fig. 3.



Excitation function for the reaction ^{6}Li($p\alpha$)^{3}He; angle of observation 135°. (*a*) ^{3}He nuclei, (*b*) ^{4}He nuclei, (*c*) ^{3}He yield corrected for motion of mass centre, (*d*) ^{4}He yield corrected for motion of mass centre, (*e*) yield of ^{7}Li($p\alpha$)^{4}He reaction on a different scale; ○ observations of Rumbaugh, Roberts and Hafstad at 90°.

and 38°, and the difference in yields for proton energies above 500 keV. suggests that the angular distribution of reaction (3) is asymmetric, though at lower energies there appears to be isotropy, in agreement with the results of Neuert (1938). The true asymmetry of the angular distribution is probably more than is suggested by fig. 3, since the curves of fig. 2, taken with a very thin target, show an even larger ^{4}He/^{3}He ratio at 135°. This is due in part to the fact that the observed yield of a group of particles of given energy increases with target thickness as previously described (II, fig. 2), and the target thickness necessary for the maximum yield of the ^{3}He particles (of energy about 2·1 MeV.) is greater than that required for the maximum yield of the ^{4}He particles (energy 1·5 MeV.). In the case of the ^{6}Li($p\alpha$) reaction the target thicknesses necessary for maximum yield of the ^{3}He and ^{4}He particles (allowing for the difference in stopping power of the target for the two particles) are about 100 and 60 μgm./cm.^{2} respectively for an analyser resolution of 4 per cent. If a rough correction for this effect is made for the 90 μgm./cm.^{2} target a ^{4}He/^{3}He ratio in agreement with that obtained from the 15 μgm./cm. target is found.

The observations of Rumbaugh, Roberts and Hafstad (1938) on the yield of this reaction are also plotted (after fitting at the 430 keV. point) in fig. 3; their figures for the ^{3}He group at 90° show a much more rapid rise than the present results, although the increase in intensity of the ^{7}Li($p\alpha$) reaction used for monitoring (fig. 3 (*e*)) was in agreement with their results. The resolution shown in the curves of fig. 1 is sufficient to make the relative excitation function reliable to about 10 per cent. As an extra check an excitation function was plotted using as detector an electron multiplier in which a mica screen was inserted to prevent the ^{4}He particles being detected. The ^{3}He particles were thus recorded with uniform efficiency over the energy range studied; the excitation curve obtained agreed well with that shown in fig. 3.

The energy release in reaction (3) was calculated from the curves of fig. 1, using "extrapolated" energies in order to eliminate the target thickness correction (*cf.* II, §2 a). A special analyser calibration for "extrapolated" energies was used to replace that of I, fig. 8, and in this calibration the Q values for the reference reactions ^{9}Be($p\alpha$)^{6}Li and ^{9}Be(pd)^{8}Be were changed from the values given in I to the latest values (2·121 MeV. and 0·558 MeV.) published by Tollestrup, Fowler and Lauritsen (1949). A carbon deposit formed on the targets during bombardment and the thickness of this film was found as described in II, §2 (a) to be 8±3 keV. for protons; a correction of 32±12 keV. to the alpha particle energy and 27±12 keV. to the ^{3}He energy has therefore been made. The results for the 90 μgm./cm.^{2} target are given in Table I.; no estimation of carbon thickness was made for the 15 μgm./cm.^{2} target, and energy releases deduced from the curves of fig. 2 have therefore not been included. The error in each of the stated Q values arises chiefly from the correction for carbon thickness and the uncertainties in proton energy (±10 keV.) and in calibration of the analyser for extrapolated energies (±15 keV.) and amounts to ±40 keV. Assuming that the observations provide three independent values for Q the mean value for the energy release in the ^{6}Li($p\alpha$) reaction is

$$Q=3{\cdot}97\pm{\cdot}03 \text{ MeV.}$$

This lies between the older values listed by Hornyak and Lauritsen (1948) and the latest figure given by Tollestrup, Fowler and Lauritsen (1949).

The cross section for reaction (3) was obtained by comparing the yield of the ^{6}Li($p\alpha$) reaction with that of the ^{10}B($p\alpha$) and ^{9}Be($p\alpha$) reactions. The ^{6}Li target used had a surface density, estimated from the ion current in the electromagnetic separator, of 5·4 μgm./cm.^{2} of the isotope before oxidation, the ^{10}B target was the 38 μgm./cm.^{2} target used in II, and the ^{9}Be target was deposited by vacuum evaporation and found to have a surface density of 16 μgm./cm.^{2} by direct weighing. The yields from these targets were assumed to be proportional to the integral $\int \frac{N(E)\,dE}{E}$ taken over the observed alpha-particle group, N(E) being the number of alpha particles recorded when the setting of the analyser corresponded to an energy E. The small variation of yield due to the slowing down of the incident protons in the target was neglected The integrals were evaluated graphically and it was found that, assuming isotropic distributions in the centre of mass system,

TABLE I.

Energy release in the ^{6}Li($p\alpha$)^{3}He reaction.

Proton energy (keV.)	^{3}He energy (keV.)	^{4}He energy (keV.)	Q(MeV.)	
			From ^{3}He energy	From ^{4}He energy
322	2122	—	3·98	—
622	2152	1482	4·00	3·95
922	2162	1482	3·96	3·95

$$\frac{\sigma^6\text{Li}\,(p\alpha)(330\text{ keV.})}{\sigma^{10}\text{B}\,(p\alpha)(330\text{ keV.})}=10{\cdot}8,$$

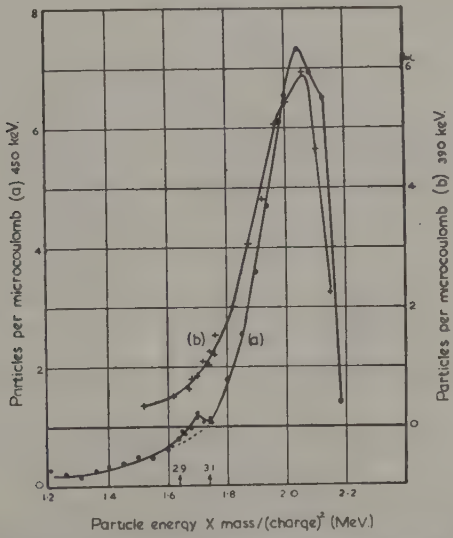
$$\frac{\sigma^6\text{Li}\,(p\alpha)(330\text{ keV.})}{\sigma^9\text{Be}\,(p\alpha)(530\text{ keV.})}=0.65.$$

The cross section for the ^{10}B($p\alpha$) reaction for a proton energy of 330 keV. was taken to be $4{\cdot}5\times10^{-27}$ cm.^{2} (II), and that for the ^{9}Be($p\alpha$) reaction for a proton energy of 530 keV. to be $1{\cdot}3\times10^{-25}$ cm.^{2} (Thomas, Rubin, Fowler and Lauritsen 1949). The corresponding values for the ^{6}Li($p\alpha$) cross section for a proton energy of 330 keV. are $4{\cdot}9\times10^{-26}$ cm.^{2} and $8{\cdot}5\times10^{-26}$ cm.^{2}. The difference between the two values is probably due to systematic errors in the cross sections assumed ; the latter value exceeds the cross section of 5×10^{-26} cm.^{2} obtained from the results of Bowersox (1939) and both are very much greater than the value of $8{\cdot}5\times10^{-27}$ cm.^{2} deduced from the curves given by Rumbaugh, Roberts and Hafstad (1938).

§4. RESULTS FOR ^{7}Li.

A search for the emission of short range alpha particles in reaction (2) was first made by bombarding a thin target of ordinary lithium with protons. Fig. 4 (a), (b) shows the results obtained with a target of surface density 17 μgm./cm.^{2} (equivalent thickness about 20 keV. at the angle used), an angle of observation of 60° and an analyser resolution of 8 per cent ; curve (a) was taken at a bombarding voltage of 450 keV., at which the ($p\gamma$) resonance was observed, and curve (b) at 390 keV. for which there should have been very little effect due to the resonance level. The corrections for target thickness and carbon deposit were made by shifting the energy scale until the ^{3}He and ^{4}He peaks fitted the values expected from the energy release in reaction (3). The position of alpha particles

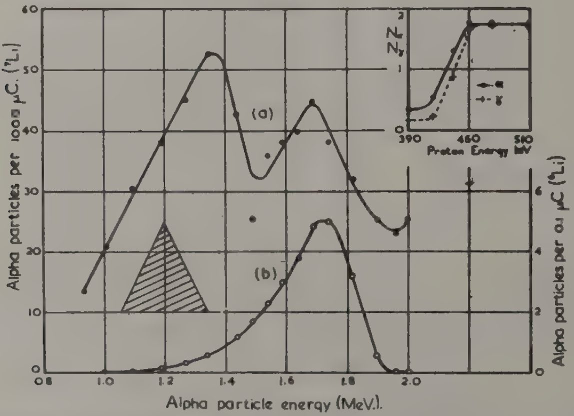
Fig. 4.



Alpha particles and ^{3}He nuclei from ordinary lithium bombarded by protons of energy (a) 450 keV. ; (b) 390 keV. The arrows show the position in which short range alpha particles emitted in the ^{7}Li($p\gamma$)^{8}Be* reaction would be expected for disintegration energies of ^{8}Be of 2900 and 3100 keV.

expected from the disintegration of a ^{8}Be nucleus with an energy release of 2900 and 3100 keV. is also shown and there is some slight evidence for the appearance of extra particles in this region on curve (*a*). The intensity of the incompletely resolved ^{6}Li($p\alpha$) group was estimated by integrating over the distribution and was compared with that of the small group which might be due to the ^{7}Li($p\gamma$)^{8}Be* reaction. Using the ^{6}Li($p\alpha$) cross section given in § 3, the thick target yield of disintegrations leading to the emission of 14·8 MeV. gamma radiation was estimated to be $2{\cdot}1 \times 10^{-10}$ per incident proton for a target of LiOH. The predicted figure is $1{\cdot}75 \times 10^{-9}$, derived from the thick target yield of $5{\cdot}35 \times 10^{-9}$ quanta (14·8+17·6 MeV. radiation) per proton (Fowler, Lauritsen and Lauritsen 1948), and the ratio 0·33 of the intensity of the 14·8 MeV. radiation to that of the total radiation at resonance (Walker and McDaniel 1948). The difference between the observed and predicted yields is probably due to the bad definition of the small alpha-particle group shown in fig. 4.

Fig. 5.



Alpha particles from lithium isotopes bombarded with 440 keV. protons. Angle of observation 104°. (*a*) ^{7}Li; (*b*) ^{6}Li. The shaded triangle gives the analyser resolution. Inset: Yield of gamma radiation and of alpha particles from ^{7}Li as a function of proton energy.

In order to reduce the effect of the ^{6}Li($p\alpha$) group, targets of the lithium 7 isotope deposited on copper backings were bombarded. The lithium 7 samples were enriched by a factor of about 500 over normal lithium, but owing to the low cross section for reaction (2) and the high cross section for the ^{6}Li($p\alpha$) process, and to the fact that the first targets used were more than 100 keV. thick, the yields from the two reactions were of the same order of magnitude. In the first experiments with thick targets, the background due to the ^{6}Li groups had to be subtracted, and the analyser resolution was therefore set at 12 per cent in order to increase the counting rate and improve the statistical accuracy. The curve given in fig. 5 (curve (*a*)) which is that already published (Burcham and Freeman 1949 a, fig. 2), but transferred to an energy scale, is typical of these results ; a bombarding energy of 440 keV. and an angle of observation of 104° were used. There is clearly a group of particles centred at an energy of $1{\cdot}38 \pm 0{\cdot}08$ MeV., and of width about 0·45 MeV., which cannot be due to contamination of the target by ^{6}Li. The group observed from a thick ^{6}Li isotopic target is shown in fig. 5 curve (*b*) ; such contamination, in the correct proportion, could evidently account for the higher energy group shown in curve (*a*). In order further to establish the identity of the group of alpha particles of mean energy 1·38 MeV., the excitation function of this group was compared with that of the total gamma radiation over the range 350–700 keV. Part of the two curves is shown in fig. 5 (inset), and the close similarity between them definitely establishes the connection between the 1·38 MeV. alpha particles and the resonant gamma radiation. With certain assumptions, discussed in §5, these results indicate the existence in ^{8}Be of a level, unstable with respect to alpha particle emission, at 2900 ± 200 keV. above the ground state and with a width of 900 keV. or less. The work described later in this section, in which the ^{6}Li contamination was less important, modifies these conclusions.

In order to see whether the higher energy group of fig. 5 (*a*) could also be attributed in part to the ^{7}Li$(p\gamma)^8$Be* reaction, the contamination of the ^{7}Li target by ^{6}Li was estimated quantitatively, using the ^{6}Li$(d\alpha)$ reaction as an indicator. The results obtained for the ^{7}Li target and a thin ^{6}Li target for comparison, were as follows :

^{7}Li target : Intensity of ^{6}Li$(d\alpha)$ group $= 65 \pm 20$ alpha particles/1000 μC.

^{6}Li target : Intensity of ^{6}Li$(d\alpha)$ group $= 203 \pm 10$ alpha particles/1000 μC.

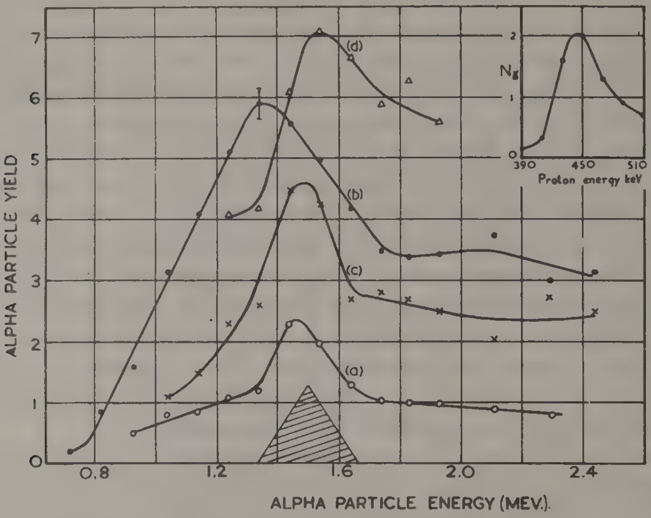
^{6}Li target : Intensity of ^{6}Li$(p\alpha)$ group $= 115 \pm 5$ alpha particles/1000 μC.

From these figures the intensity of the ^{6}Li$(p\alpha)$ group on the ^{7}Li target was expected to be 37 ± 11 per 1000 μC., while the intensity found was 44 ± 5 ; the second group could thus, within the limits of error, be ascribed entirely to ^{6}Li contamination.

For a given isotopic enrichment the ^{6}Li contamination becomes relatively less important in comparison with the effect from ^{7}Li as the target thickness decreases until a target of thickness equal to the width of the $(p\gamma)$ resonance (14 keV.) is reached. Experiments were therefore made with thinner targets ; fig. 6 shows the results obtained with an evaporated ^{7}Li$_2$SO$_4$ target. The thickness of the target is shown by the inset curve, which gives the gamma-ray yield as a function of bombarding voltage with a beam of 20 keV. spread in energy. Curves (*a*) (*b*) (*c*) (*d*) give the alpha-particle yields for bombarding voltages of 390, 450, 550 and 750 kV. taken with the analyser at an angle of 135° with the proton beam. The resonant character of the alpha-particle distribution and the presence of a group which may be ascribed to ^{6}Li contamination are again shown. The energy level in ^{8}Be deduced from these observations is centred at

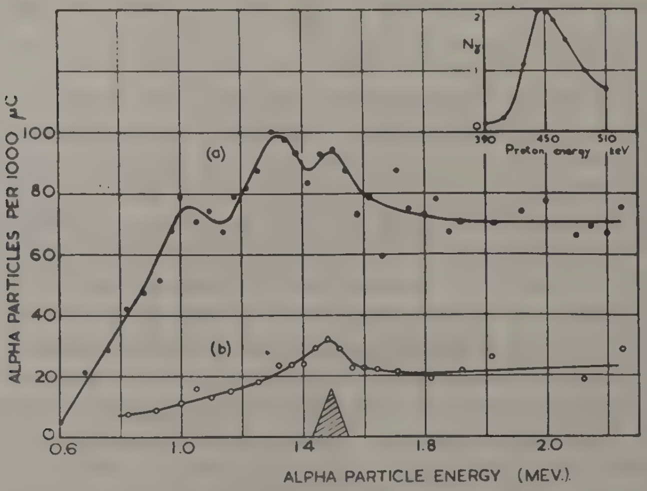
3200 keV. above the ground state, and has a width of less than 1100 keV., though both these conclusions are uncertain owing to the presence of the ^{6}Li contamination.

Fig. 6.



Alpha particles from thin $^7Li_2SO_4$ target bombarded by protons of energy (*a*) 390 keV., (*b*) 450 keV., (*c*) 550 keV., (*d*) 750 kev. Angle of observation 135°. Inset : Yield of gamma radiation as a function of proton energy.

Fig. 7.



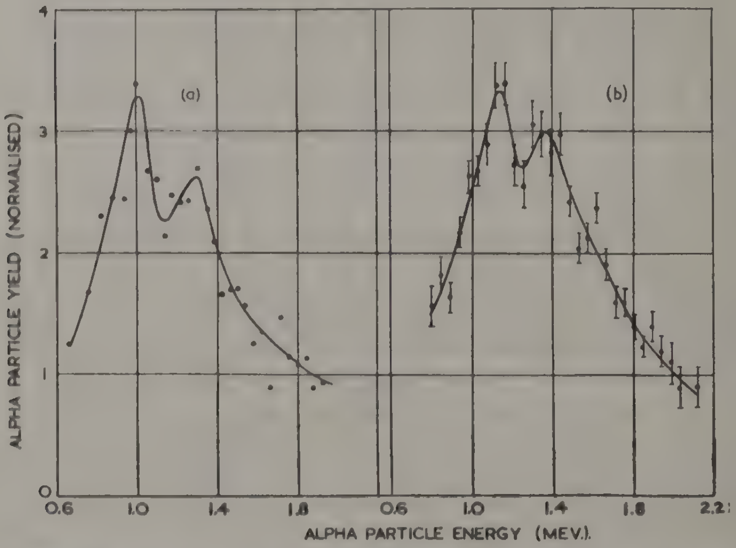
Alpha particles from a thin target containing 50μgm/cm^2 of ^{7}Li bombarded by protons of energy (*a*) 450 keV., (*b*) 390 keV. Inset: Yield of gamma radiation as a function of proton energy. The shaded triangle gives the analyser resolution.

At this stage more uniform depositis of ^{7}Li of known density became available, and it was then possible to use thin targets with a higher percentage of lithium than in Li_2SO_4. Fig. 7 gives the alpha-particle spectrum obtained from an aluminium foil on which 50 μgm./cm.^{2} of ^{7}Li had been deposited (as estimated from the ion current in the electromagnetic separator). The inset curve gives the gamma ray yield, and the alpha-particle spectra were taken for bombarding voltages of 450 keV. (curve (*a*)) and 390 keV. (curve (*b*)); in both cases the angle of observation was 135° and the analyser resolution 4 per cent in energy. The yields for the 450 keV. observations were monitored by simultaneous observation of the total gamma-ray intensity. For a target of this thickness the curves of fig. 2 and the excitation function given in fig. 3 show that the yield of particles from ^{6}Li contamination will change by only about 10 per cent between 390 and 450 keV., and the 390 keV. curve of fig. 7 has therefore been used as a background determination and subtracted from the 450 keV. alpha-particle spectrum. Since the resolution width of the analyser (60 keV. at 1500 keV.) is considerably less than the width of the alpha-particle distribution a normalizing factor for intensity must be used to give the true distribution. This factor was obtained by the method used by Thomas, Rubin, Fowler and Lauritsen (1949) in which the analyser was used to study particles scattered elastically at different energies from a thick copper target; comparison of the observed distribution with that expected from the Rutherford scattering law gives the normalizing factor.

The ordinates of fig. 7, after correction for background, were each multiplied by the observed normalizing factor, and the resulting alpha-particle distribution is plotted in fig. 8 (curve (*a*)). The shape of this distribution may still be distorted at low energy, since the normalizing factor was obtained under conditions of high gain and low bias, while in the observation of the alpha-particle spectrum the bias was increased to discriminate against scattered protons. Similar observations for proton energies of 450 keV. and 390 keV., and an angle of 135° were made with other thin targets containing 30–50 μgm./cm.^{2} of ^{7}Li; several of these targets showed no appreciable effect due to ^{6}Li contamination. Long counting periods were necessary with the analyser resolution used, and correction had to be made for the gradual growth of carbon deposit on the targets during the experiments. This growth was estimated by periodic observation of the ($p\gamma$) resonance curve. Fig. 8 (*b*) is the result of adding together all the thin target observations after correction for the carbon deposit, and after normalization and removal of background; the curve has the same appearance as that of fig. 8 (*a*), although it is shifted in energy owing to the carbon correction. Most of the thin target observations showed the structure which appears in both fig. 8 (*a*) and fig. 8 (*b*), although the statistical errors (indicated in fig. 8 (*b*)) are large, and it would be permissible to draw a curve with a single peak through the points. The interpretation of this curve is discussed in § 5; by integration over it the number of disintegrations leading to the production of 14·8 MeV. gamma radiation in the targets used (all of thickness greater than the resonance width) was found to be $1\cdot05\times10^{-9}$ per proton, which is to be compared with the predicted figure of $1\cdot75\times10^{-9}$ for a thick LiOH target. The difference is probably not significant in view of the uncertainty of target composition, and of the fact that the bias used in the alpha particle observations may have reduced the efficiency of detection below 100 per cent at low energies. The integrated yield was found to be about 20 times the peak yield with the analyser resolution used.

A simultaneous observation of the alpha particle and gamma ray excitation functions was made for one of the 30 μgm./cm.^{2} targets for alpha particles of energy 970, 1235 and 1500 keV.; the curves, of which that for 1500 keV. is shown in fig. 9, were all very similar.

Fig. 8.



Alpha particles from thin ^{7}Li targets bombarded by protons of energy 450 keV. Angle of observation 135°. (*a*) Observations shown in Fig. 7 (*a*), normalized and with background subtracted. (*b*) Combination of all normalized observations.

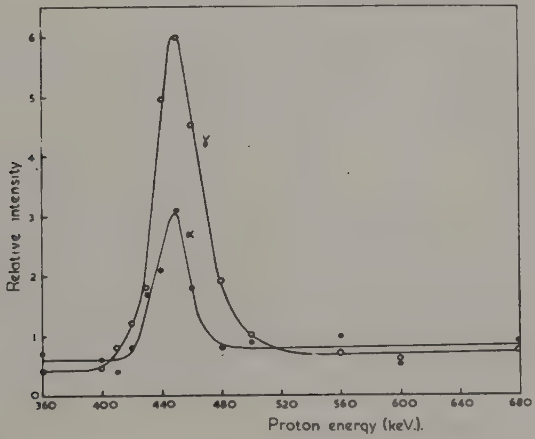
The observation of resonant alpha particles at energies as high as 1500 keV. makes it extremely unlikely that the effects observed can have been in any way due to protons multiply scattered within the analyser; the impluses from the 1500 keV. alpha particles were several times larger than those from the protons of 450 keV.

During the course of this work the alpha particle spectrum was observed at a number of different angles, chiefly with the object, in the case of the thicker targets, of finding the angle at which the separation of the ^{7}Li and ^{6}Li groups was best. Low resolution observations of this type showed no large differences in yield or spectrum of the resonant alpha particles over the range 45°–135°.

§ 5. Discussion

The results presented in § 4 make it certain that there is a resonant emission of alpha particles associated with the 14·8 MeV. resonant gamma radiation from the ^{7}Li($p\gamma$) process. Walker and McDaniel (1948) observed that the energy of the 14·8 MeV. radiation increased with increasing proton energy, and the emission of alpha particles must therefore follow that of the gamma radiation. The approximate agreement between the alpha particle and gamma ray yields makes it unlikely that there is any strongly competitive mode of break-up of the excited ^{8}Be nuclei left after the emission of the 14·8 MeV. radiation in accordance with reaction (2).

Fig. 9.



^{7}Li($p\gamma$)^{8}Be* : simulanteous excitation curves for emission of alpha particles and gamma radiation. Alpha particle energy 1500 keV.

The precise interpretation of the results given in fig. 8 (*b*) is difficult, since the alpha particles are emitted from a ^{8}Be nucleus whose momentum is partly determined by the recoil due to emission of a quantum of 14·8 MeV. radiation. The angular distribution of the 14·8 MeV. radiation with respect to the incident proton is not known in detail, and the velocity distribution of the ^{8}Be nuclei produced in reaction (2) therefore cannot be calculated. On the assumption of an isotropic distribution of recoil momenta, the single alpha particle spectrum which would be observed at 135° when the ^{8}Be nuclei are left with an excitation energy of 3000 keV. was calculated and found to be of width about 400 keV. The centre of the distribution corresponded within 50 keV. with the energy expected if the gamma ray recoil had been neglected. In interpreting the alpha particle spectra obtained in this work, it has therefore been assumed, as a first approximation, that the gamma ray momentum may be neglected for the determination of the position of the energy level involved from the observed maximum yield. The deduction of the true width of the level or levels corresponding to the observed distribution of alpha particles is more difficult, since the relative importance of the recoil effect depends on the actual width of the level. If the width due to recoil (assumed constant over the spectrum) is Δ and the true width of the level is Γ_0 the observed width should be given approximately by

$$\Gamma=\sqrt{(\Gamma_0^2+\Delta^2)}.$$

The validity of this formula was checked by graphical construction for a number of assumed values of Γ_0.

The work of Walker and McDaniel indicates a level of ^{8}Be of width about 2100 keV. Neglecting the influence of recoil, in a first approximation, this corresponds to a real width Γ_0 of 1050 keV. for the energy spectrum of single alpha particles. With $\Delta=400$ keV. the observed width would be $\Gamma=1120$ keV., which is in good agreement with the value of 1100 keV. calculated from fig. 8 (*b*). The ^{8}Be level reported by Walker and McDaniel is at an excitation of 2800 keV. above the ground state or 2900 keV. above the state representing two free alpha particles ; there is an uncertainty of about 200 keV. in this figure. Again neglecting recoil, the peak of the alpha-particle spectrum should be at 1450 ± 100 keV., which is in agreement with the value of 1560 ± 50 keV. calculated from the mid-point of the curve of fig. 8 (*b*). The present results are therefore in agreement with the work of Walker and McDaniel within the limits of error of the experiments.

The curve given in fig. 8 (*b*) shows evidence of a structure which was observed in a number of the thin target experiments made with an analyser resolution of 4 per cent. The statistical fluctuations were, however, always large, and it has not so far been possible to obtain reliable evidence for a splitting of the ^{8}Be level concerned. Further experiments, with larger intensities, and designed to display the total energy of the alpha particle pair, would be desirable to clarify this point†. The curve of fig. 8 (*b*) is therefore taken to indicate a single level in $^{8}\mathrm{Be}$ at an excitation energy of 3030 ± 100 keV. above the ground state and a real width of about 2000 keV.

---

† (*a*) R. M. Littauer and P. Meyer have recently studied the alpha-particle spectrum in this laboratory using an analyser with increased resolution, solid angle, and coverage, and with large bombarding currents. Their results suggest a single level.

(*b*) The possibility that the double peak shown in fig. 8 (*b*) is connected with a strong anisotropy in the emission of the 14·8 MeV gamma radiation is rendered unlikely by recent unpublished measurements by Devons and Lindsey on the angular distribution of this radiation. A study of the correlation between the direction of emission of the 14·8 MeV. gamma radiation and the following alpha particles is being attempted.

A level in the nucleus $^{8}\mathrm{Be}$ at an excitation energy of about 3000 keV. above the ground state has been observed in the following reactions :

$^{11}\mathrm{B}(p\alpha)^{8}\mathrm{Be}^*$	Dee and Gilbert 1936.
$^{10}\mathrm{B}(d\alpha)^{8}\mathrm{Be}^*$	Smith and Murrell 1939.
$^{4}\mathrm{He}(\alpha\alpha)^{4}\mathrm{He}$	Devons 1939.
$^{7}\mathrm{Li}(dn)^{8}\mathrm{Be}^*$	Green and Gibson 1949.
$^{12}\mathrm{C}(\gamma\alpha)^{8}\mathrm{Be}^*$	Goward, Telegdi and Wilkins 1950.

There is also evidence of a similar level in $^{8}\mathrm{Be}$ at nearly the same excitation, which takes part in the break-up of $^{8}\mathrm{Li}$; this has been established by observations on the decay of $^{8}\mathrm{Li}$ produced in the following reactions :

$^{7}\mathrm{Li}(dp)^{8}\mathrm{Li}$	Christy, Cohen, Fowler, Lauritsen and Lauritsen 1947 ; Bonner, Evans, Malich and Risser 1948 ; Baxter, Burcham and Paul 1950.
$^{11}\mathrm{B}(n\alpha)^{8}\mathrm{Li}$	Lawrance 1939.
$^{9}\mathrm{Be}(\gamma p)^{8}\mathrm{Li}$	Titterton 1950.

None of these reactions, with the possible exception of $^{7}\mathrm{Li}(dn)$, suggests a level width of as much as 2000 keV.; most of them are consistent with a width of about 800 keV., and it is therefore not at present possible to identify the level concerned in the $^{7}\mathrm{Li}(p\gamma)^{8}\mathrm{Be}^*$ reaction with any previously known level of $^{8}\mathrm{Be}$. It is, however, certain that this level lies within about 200 keV. of those deduced from other experiments. In the particular case of the break-up of $^{8}\mathrm{Li}$, Baxter, Burcham and Paul (1950) using the same analyser as in the present experiments, found a level at 3130 ± 50 keV., which is to be compared with the value of 3030 ± 100 keV. found in the present work.

The excitation curves given in fig. 9 suggest that the alpha particle yield at resonance increases less sharply than that of the total gamma radiation. This might be due to an uncertainty in the background levels owing to the low counting rates, but it would be in agreement with the observation of Devons and Hine (1949) that the ratio of the 14·8 to the 17·6 MeV. gamma radiation decreases at the 440 keV. proton resonance energy.

ACKNOWLEDGMENTS

We are grateful to Sir John Cockcroft and to Dr. W. D. Allen and members of his group at the Atomic Energy Research Establishment, Harwell, who provided the separated lithium isotopes. We also wish to thank Messrs W. Birtwhistle and D. D. Stewart for much technical assistance.

REFERENCES

BAXTER, A. S., BURCHAM, W. E., and PAUL, E. B., 1950, *Phil. Mag.* (to be published).

BONNER, T. W., and EVANS, J. E., 1948, *Phys. Rev.*, **73**, 666.

BONNER, T. W., EVANS, J. E., MALICH, C. W., and RISSER, J. R., 1948, *Phys. Rev.*, **73**, 885.

BOWERSOX, R. B. 1939, *Phys., Rev.*, **55**, 323.

BURCHAM, W. E., and FREEMAN, J. M., 1949 a, *Phys. Rev.*, **75**, 1756, **77**, 287; 1949 b, *Phil. Mag.*, **40**, 807 ; 1950, *Ibid.*, **41**, 337.

CHRISTY, R. F., COHEN, E. R., FOWLER, W. A., LAURITSEN, C. C., and LAURITSEN, T., 1947, *Phys. Rev.*, **72**, 698.

DEE, P. I., and GILBERT, C. W., 1936, *Proc. Roy. Soc.*, A, **154**, 279.

DEVONS, S., 1939, *Proc. Roy. Soc.*, A, **172**, 564.

DEVONS, S., and HINE, M. G. N. 1949, *Proc. Roy. Soc.*, A, **199**, 56.

FOWLER, W. A., LAURITSEN, C. C., and LAURITSEN, T., 1948, *Rev. Mod. Phys.*, **20**, 236.

GOWARD, F. K., TELEGDI, V. and WILKINS, F. J., 1950. *Proc. Phys. Soc.*, A, **63**, 402.

GREEN, L. L., and GIBSON, W. M., 1949, *Proc. Phys. Soc.*, **62**, 407

HORNYAK, W. F., and LAURITSEN, T., 1948, *Rev. Mod. Phys.*, **20**, 191.

LAWRANCE, A. M., 1939, *Proc. Camb. Phil. Soc.*, **35**, 304.

NEUERT, H., 1938, *Phys. Zeits.*, **38**, 618.

RUMBAUGH, L. H., ROBERTS, R. B., and HAFSTAD, L. R., 1938, *Phys. Rev.*, **54**, 654.

SMITH, C. L., and MURRELL, E. B. M., 1939, *Proc. Camb. Phil. Soc.*, **35**, 298.

TANGEN, R., 1946, *Kgl. Norske Skrifter* No. 1.

THOMAS, R. G., RUBIN, S., FOWLER, W. A., and LAURITSEN, C. C., 1949, *Phys. Rev.*, **75**, 1612.

TITTERTON, E. W., 1950, *Nature, London.*, **165**, 721.

TOLLESTRUP, A. V., FOWLER, W. A., and LAURITSEN, C. C., 1949, *Phys. Rev.*, **76**, 428.

WALKER, R. L., and MCDANIEL, B. D., 1948, *Phys. Rev.*, **74**, 315.

WARSHAW, S. D., 1949, *Phys. Rev.*, **76**, 1759.

WHEELER, J. A., 1941, *Phys. Rev.*, **59**, 27.

LXXXV. *Short Range Alpha Particles from the Break-up of* ^{8}Li.

By A. S. Baxter, W. E. Burcham and E. B. Paul[†],
Cavendish Laboratory, Cambridge[‡].

[Received June 27, 1950.]

Summary.

The energy spectrum of the alpha particles from the ^{8}Li–^{8}Be* break-up has been studied using a magnetic analyser and thin targets. The most probable energy of the alpha particle pair emitted was found to be 3150±50 keV.

§1. Introduction.

The nucleus ^{8}Li is an electron emitter of 0·89 sec. half-life which breaks up in accordance with the reaction

$$^{8}\mathrm{Li} \rightarrow {}^{8}\mathrm{Be}^{*}+\beta^{-} \rightarrow 2\,{}^{4}\mathrm{He}+\beta^{-} \qquad \ldots\ldots\ldots \quad (1)$$

with an energy release of 16·75 MeV. (Hornyak and Lauritsen 1950). The spectrum of alpha particles emitted in this process has been studied be several workers and an analysis of their results up to 1941 is given by Wheeler (1941 a). The main conclusion drawn by Wheeler is that the distribution of alpha particles at low energies can be accounted for by supposing the existence of a broad level in the ^{8}Be nucleus at an excitation energy of about 3 MeV.; reasons are given for believing this level to be distinct from another level in ^{8}Be at about the same excitation which is observed in the ^{11}B($p\alpha$) reaction (Dee and Gilbert 1936) and in the scattering of alpha particles by helium (Devons 1939, Wheeler 1941 b). Recently new observations have been made on the energy distribution of both the alpha particles and beta particles emitted in reaction (1); Christy, Cohen, Fowler, Lauritsen and Lauritsen (1947) studied the momentum relations in the process and Bonner, Evans, Malich and Risser (1948) extended observations of the alpha particle spectrum to very low energies by projecting ^{8}Li nuclei into a cloud chamber. Cassels has obtained similar results, with better statistical accuracy, by projecting ^{8}Li nuclei into a photographic plate[§]. Hornyak and Lauritsen (1950) showed that existing data on the low energy alpha particle spectrum fitted with their measurements of the beta particle energy distribution, and that less than 2 per cent of all the β^- transitions from ^{8}Li proceed to the ground state of ^{8}Be.

[†] Now at Chalk River Laboratories, Ontario, Canada.

[‡] Communicated by the Authors.

[§] We are grateful to Dr. Cassels for informing us of his results before publication.

The present work was commenced early in 1948 with the object of improving the statistical accuracy of the existing observations of the low energy part of the alpha particle spectrum. The method of magnetic analysis used, in which the energy of only one member of the alpha particle pair from reaction (1) is observed, is inferior to methods which display the total energy release in each disintegration, and leads, as pointed out in § 3, to a broadening of the spectrum. It is felt, however, that the present results are of some interest in showing the application of a magnetic analyser to a study of the ^{8}Li reaction.

§2. EXPERIMENTAL.

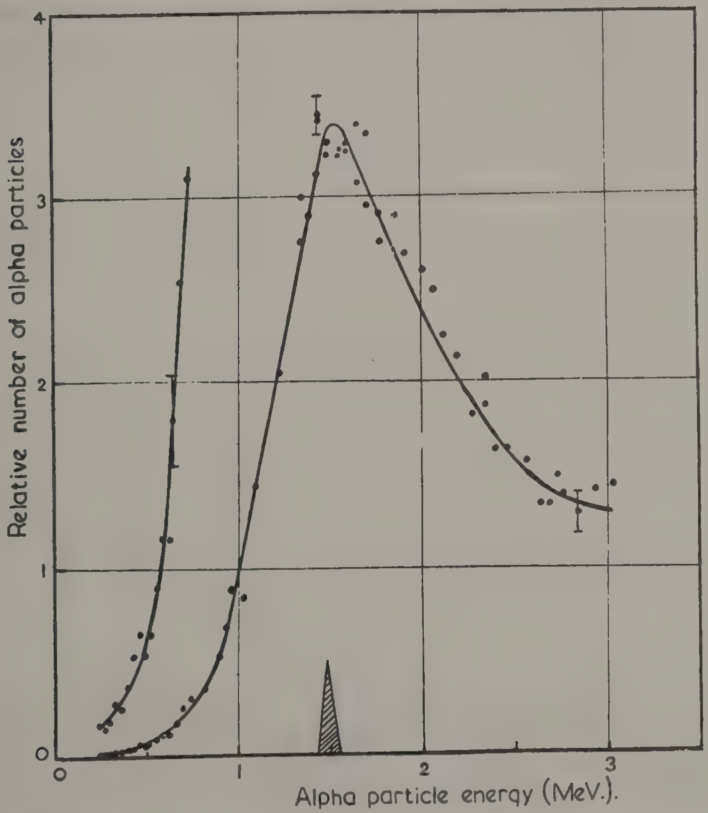
The alpha particle spectrum was investigated by means of the magnetic analyser described in other papers (Burcham and Freeman 1949, 1950 a, 1950 b). In the first experiments a lithium hydroxide target deposited on a thick copper backing was bombarded with deuterons of 900 keV. energy. The energy release in the reaction ^{7}Li(dp)^{8}Li is $-0{\cdot}193\pm0{\cdot}008$ MeV. (Strait and Buechner 1949, Paul 1950) and the ^{8}Li nuclei are all projected forwards, the maximum angle of recoil being 34° with the deuteron beam. The ^{8}Li nuclei, therefore, recoil into the target and target backing and a correction, amounting to about 100 keV. had to be made in consequence to the observed energy of the alpha particles emitted (*cf.* Bonner, Evans, Malich and Risser 1948). This correction is entirely eliminated in the cloud chamber and photographic plate experiments; in later work with the magnetic analyser it was reduced by using thin foils of collodion as target backing.

The collodion foil on which the lithium hydroxide was deposited was set at 34° with the deuteron beam and the radioactive alpha particles were observed in a direction at right angles to the foil. The thickness of both the foil and the target were measured by observing the 339 keV. ^{19}F($p\gamma$) resonance with and without the foil in position, and were found to be 20 ± 5 keV. and 10 ± 5 keV. respectively for protons. The observations were made using a suitable irradiation and counting cycle; during the 2 second irradiation period the yield of the reaction was monitored by observation of the alpha particles from the ^{6}Li($d\alpha$) reaction through a side window, and the radioactive alpha particles were counted, for different analyser settings, in a 2 second period following the interruption of the deuteron beam. It was found possible to bombard the collodion foils with currents of about 5 μA. without damage for a period of about an hour, and as the yield was large (about 400 radioactive alpha particles per second at the peak of the spectrum) good statistical accuracy was possible. The targets became blackened by deposition of carbon during the experiment, but repeat observations showed only a small shift of the spectrum over a period of an hour. The analyser resolution was 4 per cent in energy.

§3. Results.

The alpha particle spectrum obtained is shown in fig. 1, in which the number of alpha particles for a given number of irradiation cycles, corrected by using the monitor counts to the same intensity of irradiation, is plotted against energy. The lower energy portion of the spectrum is shown on a larger scale. The total thickness of the backing foil and target was 30±10 keV. for protons of 340 keV. energy, or approximately 120±40 keV. for alpha particles of about 1500 keV., and in order to correct for this it was assumed that the average retardation of the alpha particles would be 60±20 keV. Correction was also made for the varying transmission factor of the analyser over the spectrum by applying an experimentally determined normalizing factor (*cf.* Burcham and Freeman 1950 b). The effect of charge exchange on the low energy particles was also allowed for by using the figures given by Briggs (1927) ; the ordinates at the very

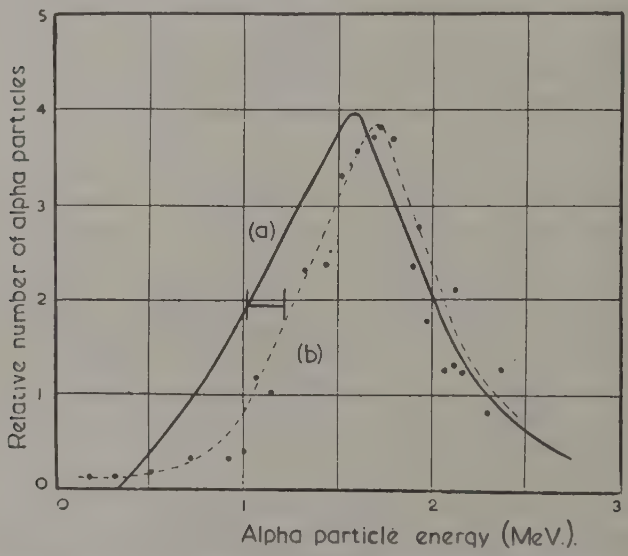
Fig. 1.



Alpha particles from the ^{8}Li–^{8}Be* break-up ; experimental points. The shaded triangle represents the analyser resolution.

low energies have been increased by the necessary factor and the corresponding high energy points reduced to allow for detection of singly charged particles. With these corrections and normalization the spectrum of fig. 2 (*a*) is obtained from the smoothed curve of fig. 1 ; this figure shows that the maximum yield occurs for an energy of 1575±25 keV. If the alpha particles are assumed to be emitted from ^{8}Be nuclei at rest the energy of the most probable alpha particle pair is 3150±50 keV. which is to be compared with the value of 3300 keV. found by Bonner, Evans, Malich and Risser (1948). The points obtained by these authors are plotted for comparison in fig. 2, with the assumption that the energy of the alpha particle pair divides equally between the two alpha particles, and it is clear that the width (about 1000 keV.) of the present distribution is considerably greater. This may be due to the corrections applied to the experimental points, although the increased width must in part be due to the use of a target of thickness 120 keV. as a source of activity, and to the fact that the ^{8}Be* nucleus disintegrates into alpha paticles before it has lost its energy of recoil from the emission of the electron from ^{8}Li, so that the two alpha particles in general have different energy. The recoil effect depends on the distribution of recoil momentum between the ^{8}Be* nucleus and neutrino, and has been studied by Christy, Cohen, Fowler, Lauritsen and Lauritsen (1947). With the assumption of the maximum

Fig. 2.



Alpha particles from the break-up of ^{8}Li ; corrected distribution in energy. (*a*) present observations, (*b*) experimental points found by Bonner, Evans, Malich and Risser (1948). The horizontal line shows the broadening expected in curve (*a*) as a result of target thickness and recoil.

recoil energy of about 10 keV. and an isotropic distribution of recoiling ^{8}Be* nuclei in a target of 120 keV. thickness, it was found by construction that a spectrum of the form given by Bonner, Evans, Malich and Risser (1948) (theoretical curve) would be broadened by about 200 keV. in E_α without appreciable shift of the peak. The recoil effect is responsible for about 125 keV. of this broadening and the target thickness for about 75 keV. The broadening is indicated in fig. 2, and correction for it brings the width of the present distribution into agreement with the cloud chamber work, although it makes it unsuitable for detailed comparison with a theoretical level formula.

In deducing the position of the excited state of ^{8}Be* effective in the break-up of ^{8}Li, allowance must be made for the energy dependent factor $(Q—2E_\alpha)^5$ $(Q=16{\cdot}75$ MeV.) which appears in the theoretical formula for the yield of alpha particles pairs of energy $2E_\alpha$ to take account of the varying probability of the β-transition (Wheeler 1941). If the ordinates of the distribution given in fig. 2 are divided by this factor the peak is found to shift to 1610 ± 25 keV., which corresponds to an energy for the excited state of 3130 ± 50 keV., using the figure of 89 ± 5 keV. for the energy of the ground state of ^{8}Be given by Tollestrup, Fowler and Lauritsen (1949).

ACKNOWLEDGMENTS.

We wish to thank Messrs. W. Birtwhistle and D. D. Stewart for much technical assistance.

REFERENCES.

BONNER, T. W., EVANS, J. E., MALICH, C. W. and RISSER, J. R., 1948, *Phys. Rev.*, **73**, 885.

BRIGGS, G. H., 1927, *Proc. Roy. Soc.* A, **114**, 341.

BURCHAM, W. E. and FREEMAN, J. M., 1949, *Phil. Mag.*, **40**, 807.

BURCHAM, W. E. and FREEMAN, J. M., 1949, *Phil. Mag.*, **40**, 807; 1950 a, *Ibid.*, **41**, 337; 1950 b, *Ibid.*, **41**, 921.

CHRISTY, R. F., COHEN, E. R., FOWLER W. A., LAURITSEN, C. C. and LAURITSEN, T., 1947, *Phys. Rev.*, **72**, 698.

DEE, P. I. and GILBERT, C. W., 1936, *Proc. Roy. Soc.* A, **154**, 279.

DEVONS, S., 1939, *Proc. Roy. Soc.* A, **172**, 564.

HORNYAK, W. F. and LAURITSEN, T., 1950, *Phys. Rev.*, **77**, 160.

PAUL, E. B., 1950, *Phil. Mag.*, **41**, 942.

STRAIT, E. N. and BUECHNER, W. W., 1949, *Phys. Rev.*, **76**, 1766.

TOLLESTRUP, A. V., FOWLER, W. A. and LAURITSEN, C. C., 1949, *Phys. Rev.*, **76**, 428.

WHEELER, J. A., 1941 a, *Phys. Rev.*, **59**, 27; 1941 b, *Ibid.*, **59**, 16.

LXXXVI. *The Energy Release in the Reaction* ^{7}Li(dp)^{8}Li.

By E. B. Paul*,
Cavendish Laboratory, Cambridge†.

[Received June 27, 1950.]

SUMMARY.

In the reaction ^{7}Li(dp)^{8}Li the velocity of the centre of mass exceeds the velocity of the recoil nucleus in the centre of mass system and recoils will be observed only at angles less than a certain angle with the deuteron beam. This maximum angle has been observed for several deuteron energies and the Q-value for the reaction determined to be $-0\cdot187\pm0\cdot010$ MeV.

§1. INTRODUCTION.

THE energy release in the reaction ^{7}Li(dp)^{8}Li has been determined by several methods. Rumbaugh, Roberts and Hafstad (1938) using a target enriched in the ^{7}Li isotope observed no protons of range greater than 1·76 cm. with 0·86 MeV. deuterons. This sets an upper limit to the Q-value of 0·28 MeV. They also found no appreciable yield of ^{8}Li at deuteron energies less than 0·36 MeV. which sets a lower limit to the Q-value of $-0\cdot20$ MeV. Strait and Buechner (1949) observed a proton group from bombardment of lithium with 1·36 MeV. deuterons. The protons were analysed magnetically and detected in photographic plates; the Q-value reported for the ^{7}Li(dp)^{8}Li reaction is $-0\cdot193\pm0\cdot008$ MeV.

The present paper describes work done at the Cavendish Laboratory in 1948 employing a different method of finding the Q-value for this reaction, based on the observation of the maximum angle of recoil of the ^{8}Li nuclei. This method, which can be used to study a reaction for which the velocity of the centre of mass exceeds that of the recoil nucleus in the centre of mass system, was suggested by Newson (1935) and has been discussed by Bethe (1937) who shows that the maximum angle of recoil ϕ_0 in the laboratory system is given by

$$\sin\phi_0=\left\{\frac{M_2}{M_1M_3}\left[(M_0+M_1)\frac{Q}{E_1}+M_0\right]\right\}^{1/2}, \quad . \quad . \quad . \quad . \quad (1)$$

where M_0, M_1, M_2, M_3 are the masses of the initial nucleus and particle, and of the final particle and nucleus respectively and E_1 is the energy of the incident particle. For the ^{7}Li(dp)^{8}Li reaction the condition that ϕ_0

---

* Now at Chalk River Laboratories, Chalk River, Ontario.

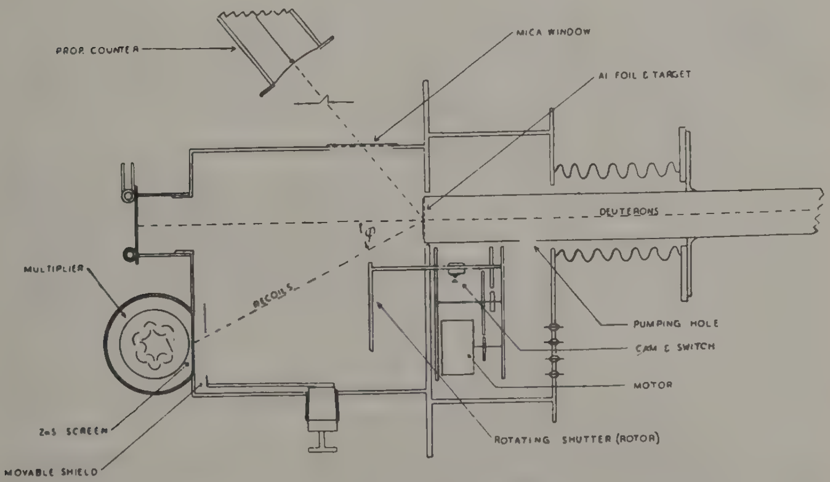
† Communicated by Dr. W. E. Burcham.

shall be real is that $E_1 > |Q|$ approximately, so that a maximum angle of recoil should be observable for deuteron energies exceeding 0·2 MeV. An examination of relation (1) shows that the sensitivity of the method is greatest when E_1 is as small as possible and it may be shown that the accuracy of the Q-value so determined is least dependent on E_1 when Q is near zero. The number of recoils per unit solid angle reaches a maximum at the angle ϕ_0 for an isotropic distribution of recoils in the centre of mass system and this facilitates observation of ϕ_0.

§2. Experimental.

The experimental arrangement is shown in fig. 1. The deuteron beam passed through a grid which supported an aluminium foil approximately 0·61 mgm./cm.^{2} in thickness. A thin film of lithium oxide was deposited on the foil. ^{8}Li recoil nuclei produced in the target by deuteron bombardment were received on a zinc sulphide screen set at an angle near the maximum expected. A movable shield in front of the fluorescent screen varied the angular range over which recoils could reach the screen from the target. The alpha particles from the break-up of the ^{8}Li nuclei received by the fluorescent screen produced scintillations which were detected by an RCA 931A electron multiplier placed close behind the screen.

Fig 1.



Apparatus used for detection of ^{8}Li recoils produced in the ^{7}Li(dp)^{8}Li reaction.

Periodic irradiations of 2·3 sec. were made and the alpha-particle activity on the screen was observed in periods of 3·2 sec. following each irradiation. During these counting periods the deuteron beam was interrupted and a rotating shutter automatically moved between the target and the screen in order to prevent detection of any alpha particles directly from the target.

The angle fixed by the edge of the shield was measured by replacing the target by a sheet of tracing paper illuminated from behind. The distance from the point at which the beam intersected the plane of the fluorescent screen to the edge of the shadow on the screen was measured with dividers.

The movable shield allowed ^{8}Li recoils to be recorded at angles greater than that fixed by the edge of the shield but the exposed screen area was different for each setting of the shield. This correction was found by placing a polonium source in the target position and counting for different settings of the shield. The correction curve so obtained was used to allow for the change in total solid angle with angle of recoil.

The air equivalent of the target material and supporting foil was measured before and after bombardment using a polonium source and an ionization chamber. The difference between the residual extrapolated range of the polonium alpha particles when the target foil was in place and removed was $0{\cdot}81 \pm 0{\cdot}04$ cm. air.

A proportional counter viewed the target continuously through a mica window in the side of the vacuum box. The counter was arranged to detect the alpha particles from the reaction $^6\text{Li}(d\alpha)^4\text{He}$ and served to monitor the conditions of irradiation.

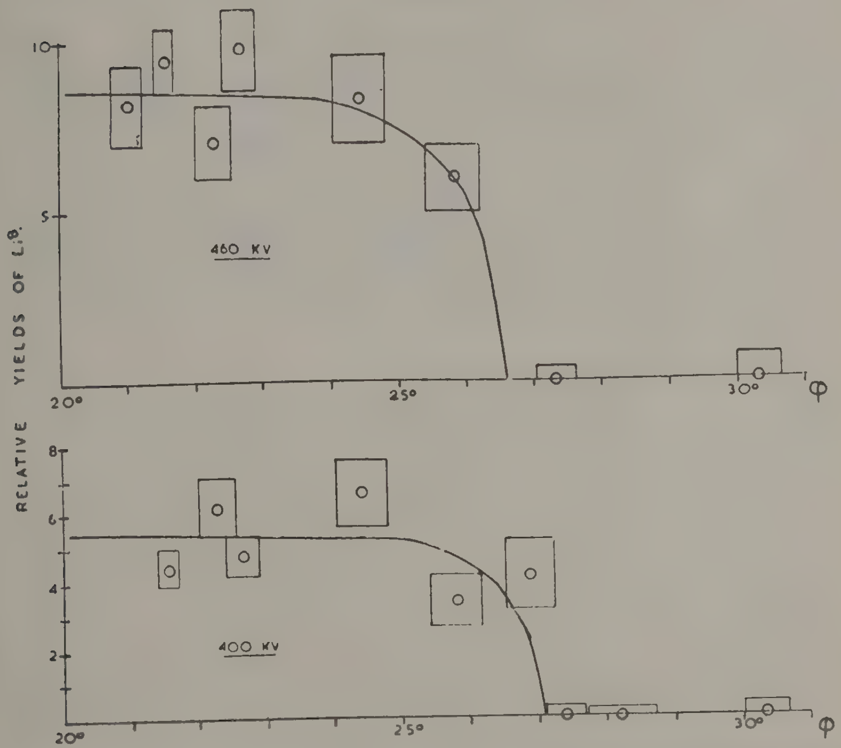
About 40 irradiations were performed for each setting of the movable shield. The background due to dark current in the multiplier and gamma radiation from the target was measured when the shield completely covered the screen. Runs were made with primary deuteron energies of 950, 900 and 850 keV. The energies of these deuteron beams after having passed through the aluminium foil and lithium oxide target were taken to be 550 ± 20 keV., 460 ± 30 keV., and 400 ± 30 keV. respectively. ^{8}Li nuclei could escape from the target from a depth of about 0·4 mm. air equivalent, that is, about 10 per cent of the thickness of the lithium target. Thus, for example, deuterons of initial energy 950 keV. produced ^{8}Li nuclei which emerged from the target only for a deuteron energy of between 570 and 530 keV. This was assumed to give the uncertainty in E_1 since other sources of error were estimated to be less important. Scattering of the ^{8}Li nuclei before emerging from the target was neglected, although such scattering probably contributed to the uncertainty in the measurement of the maximum angle of recoil.

The background count per irradiation was subtracted from the total count per irradiation and correction was also made for variations in the count from the monitor counter. Finally the count of delayed alpha particles for each angle setting was multiplied by the corresponding geometrical factor as determined with a polonium source in place of the target. This gave the mean count per unit solid angle per irradiation for each setting of the shield. A check made of the period of the activity being observed gave a half-life of $1{\cdot}3 \pm 0{\cdot}5$ sec.

§3. RESULTS.

The corrected results of the various runs are presented in figs. 2 and 3. The ordinates correspond to counts per irradiation for 100 counts per irradiation in the monitor counter for a solid angle equal to that subtended by the screen in the fully exposed position. The values of the maximum angle are listed in the Table together with the Q-value calculated for the reaction.

Fig. 2.

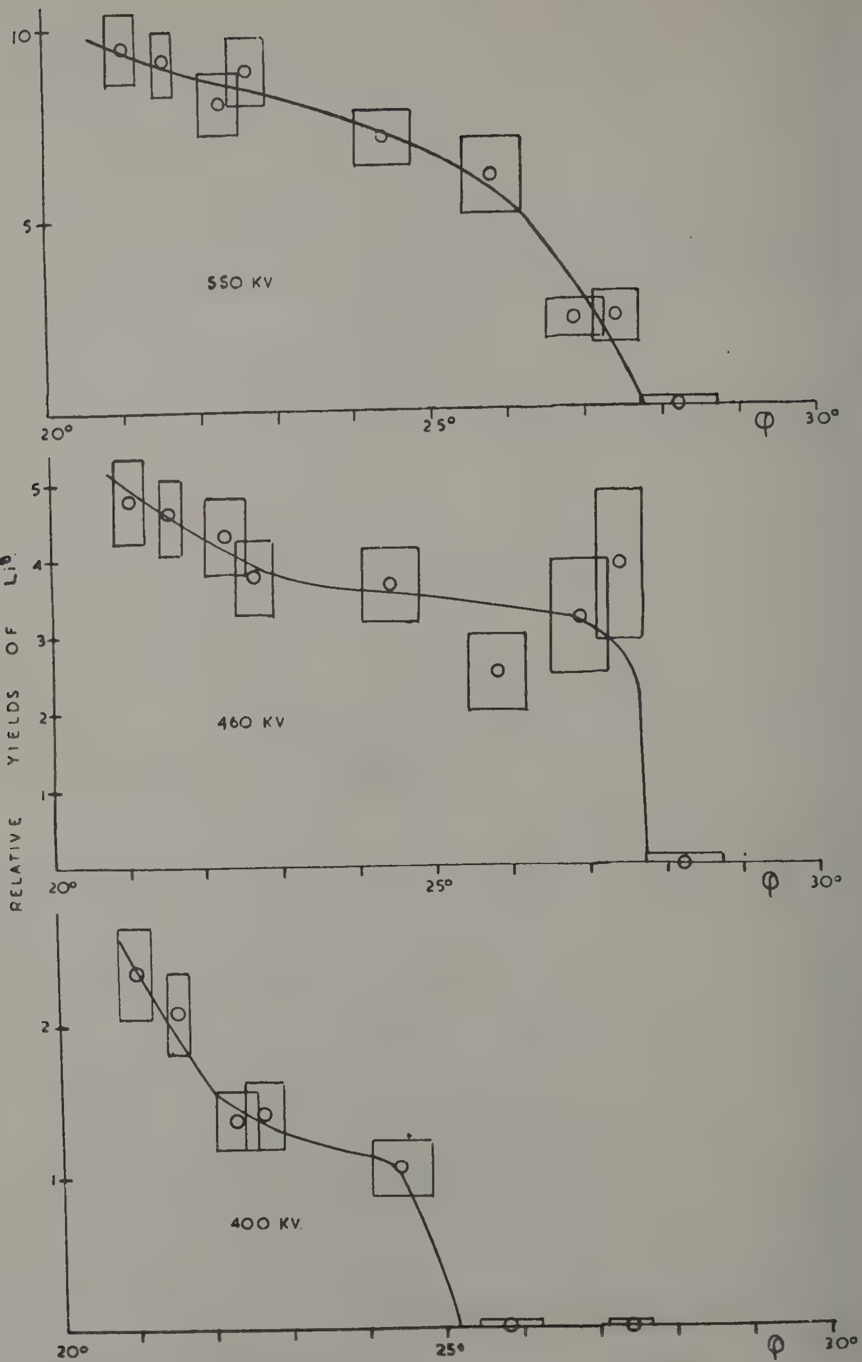


Angular distribution of recoils from the ^{7}Li(dp)^{8}Li reaction. Runs 1 and 2.

TABLE.

Run	E_1 (keV.)	ϕ_0 (degrees)	Q (MeV.)
1	460±30	26·6±1·5	−0·193±0·024
2	400±30	27·2±0·5	−0·162±0·013
3	550±20	27·8±0·8	−0·215±0·014
4	460±30	27·7±0·8	−0·182±0·015
5	400±30	25·2±1·4	−0·182±0·021

Fig. 3.



Angular distribution of recoils from the ^{7}Li(dp)^{8}Li reaction. Runs 3, 4 and 5

These results have been combined and the weighted average for the energy release in the reaction ^{7}Li(dp)^{8}Li is found to be

$$Q=-0{\cdot}187\pm0{\cdot}010 \text{ MeV.}$$

This value is in satisfactory agreement with the values of $-0{\cdot}20\pm0{\cdot}03$ MeV. found by Rumbaugh, Roberts and Hafstad (1938) and of $-0{\cdot}193\pm0{\cdot}008$ found by Strait and Buechner (1949).

ACKNOWLEDGMENTS.

The writer is grateful to Dr. W. E. Burcham for advice in the course of this experiment and to Mr. W. Birtwhistle for technical assistance.

REFERENCES.

BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 281.
NEWSON, H. W., 1935, *Phys. Rev.*, **48**, 790.
RUMBAUGH, L., ROBERTS, R. and HAFSTAD, L. R., 1938, *Phys. Rev.*, **54**, 657.
STRAIT, E. N. and BUECHNER, W. W., 1949, *Phys. Rev.*, **76**, 1766.

LXXXVII. *On the Quantum Mechanics of Fluids.*

By P. J. Price,

Royal Society Mond Laboratory, Cambridge*.

[Received July 10, 1950.]

Abstract.

The position of this paper, in the quantum theory of liquids, is principally that of a critique of the contribution by Green. The relation of his formalism to wave-mechanics is discussed, and hydrodynamical equations are derived, in wavemechanical form, which are shown to be equivalent to Green's results. A formula for the boundary stress is arrived at, and it is shown that the difference between the kinetic and thermodynamic pressures—on which Green's explanation of the thermomechanical effects in liquid helium II is based—is identically zero, confirming the conclusions of Yvon and de Boer.

§1. Introduction.

The system discussed here is that of N identical atoms (bosons or fermions), without effective internal degrees of freedom or spin, enclosed in a volume V and interacting by a potential U. H. S. Green (Born and Green 1947, Green 1948: referred to below as I and II respectively) has given the most complete account to date of the quantum mechanics of this model, and derived the hydrodynamical equations of motion in a suggestive form which demonstrates their correspondence to the classical equations. In this paper a treatment by conventional wavemechanical methods is given, and the results shown to be equivalent to Green's. The virial theorem is derived, and the stress on the boundary identified in the surface term. A discussion of the thermodynamic pressure follows, in which de Boer's (1949) conclusion is confirmed: that Green's expression for the difference between the kinetic and thermodynamic pressures vanishes identically.

§ 2. Green's Formalism and Wave-Mechanics.

We need a means of comparing results in Green's formalism (I, § 2) with those derived below. For simplicity, consider a single particle only, with wave-function $\psi(\mathbf{r})$ defined in a domain $\mathfrak{R}$ and satisfying a boundary condition B on its boundary, S. If $\{\phi_\nu\}$, $(\nu=1, 2, \ldots)$ is a complete orthonormal set of such functions, then

$$\delta(\mathbf{r}_1, \mathbf{r}_2) \equiv \sum_{\nu=1}^{\infty} \phi_\nu(\mathbf{r}_1)\phi_\nu^*(\mathbf{r}_2) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

---

* Communicated by the Author.

has the property

$$\int^{\mathfrak{R}} \delta(\mathbf{r}_1, \mathbf{r}_2)\psi(\mathbf{r}_2)d^3\mathbf{r}_2 = \psi(\mathbf{r}_1) \text{ in } \mathfrak{R},$$

and obeys B on S.

In Dirac's notation, if ψ_k is the wave-function corresponding to a ket $|\,k\rangle$,

$$\psi_k(\mathbf{r}') \equiv \langle \mathbf{r}' \,|\, k\rangle.$$

Hence by (1) $$\langle \mathbf{r}' \,|\, \mathbf{r}''\rangle = \delta(\mathbf{r}', \mathbf{r}''). \qquad (2)$$

If $$\Delta_1 \equiv \Delta(\mathbf{r}_1) \equiv |\,\mathbf{r}_1\rangle\langle \mathbf{r}_1\,|, \qquad (3)$$

then by (2), $$\Delta_1 = \delta(\mathbf{r}_1, \mathbf{r}), \qquad (4)$$

where $\mathbf{r}$ is a q-number. An operator A is expressed in the Schrödinger notation by

$$\langle \mathbf{r}' \,|\, \mathrm{A} \,|\, k\rangle \equiv \mathrm{A}\psi_k(\mathbf{r}'), \qquad (5)$$

and in Green's formalism by

$$\langle \mathbf{r}' \,|\, \mathrm{A} \,|\, \mathbf{r}''\rangle \equiv \mathrm{A}(\mathbf{r}', \mathbf{r}''). \qquad (6)$$

Now if $|\,k\rangle\langle k\,| \equiv \rho_k$ is the projection operator of a pure state k,

$$\rho = \sum_k a_k \rho_k$$

(where the a_k are real and $\geqslant 0$) is von Neumann's statistical operator (von Neumann 1943, Dirac 1947). By (5) and (6), if A and B are hermitian,

$$\begin{aligned}[\mathrm{A}\rho_k\mathrm{B}](\mathbf{r}', \mathbf{r}') &= \langle \mathbf{r}' \,|\, \mathrm{A}\rho_k\mathrm{B} \,|\, \mathbf{r}'\rangle \\ &= \langle k \,|\, \mathrm{A}\Delta(\mathbf{r}')\mathrm{B} \,|\, k\rangle \\ &= [\mathrm{A}\psi_k(\mathbf{r}')]^*\mathrm{B}\psi_k(\mathbf{r}'). \qquad (7\end{aligned}$$

Hence $$\left.\begin{aligned} &\left[\sum_\lambda \mathrm{A}_\lambda\rho\mathrm{B}_\lambda\right](\mathbf{r}', \mathbf{r}') = 0 \qquad (a) \\ \text{implies} \\ &\sum_\lambda [\mathrm{A}_\lambda\psi(\mathbf{r}')]^*\mathrm{B}_\lambda\psi(\mathbf{r}') = 0 \qquad (b) \end{aligned}\right\} \qquad (8)$$

and conversely.

All Green's final c-number equations are in the form (8 a) (or an N-particle generalization, involving summations and integrations over the 3N coordinates), and hence have equivalent wave-mechanical forms (8 b). For example, Green's

$$\{\rho, \mathbf{p}\} \equiv \tfrac{1}{2}(\rho\mathbf{p} + \mathbf{p}\rho)$$

corresponds to Landau's (1941) current operator

$$\mathbf{J}(\mathbf{r}') \equiv \tfrac{1}{2}(\Delta(\mathbf{r}')\mathbf{p} + \mathbf{p}\Delta(\mathbf{r}')),$$

whose expectation

$$\langle k \,|\, \mathbf{J}(\mathbf{r}') \,|\, k\rangle = \tfrac{1}{2}i\hbar[\psi_k\nabla\psi_k^* - \psi_k^*\nabla\psi_k](\mathbf{r}'),$$

is the conventional current at the point $\mathbf{r}'$. Similarly

$$\left(\frac{1}{m}\right)\{\{\rho, \mathbf{p}\}, \mathbf{p}\} \rightarrow -\frac{\hbar^2}{4m}[(\psi^*\nabla\nabla\psi + \psi\nabla\nabla\psi^*) - (\nabla\psi^*\nabla\psi + \nabla\psi\nabla\psi^*)]. \qquad (9)$$

(The arrow denotes the correspondence (8), and the equality—for a pure state—(7)). Each of (8) is equivalent to

$$\mathrm{Tr}\left[\rho\sum_\lambda A_\lambda \Delta(\mathbf{r}')B_\lambda\right]=0,$$

and hence to the *classical* equation

$$\iint \rho(\mathbf{p},\mathbf{r})\sum_\lambda A_\lambda(\mathbf{p},\mathbf{r})B_\lambda(\mathbf{p},\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')\,d^3\mathbf{p}\,d^3\mathbf{r}$$

$$=\int \rho(\mathbf{p},\mathbf{r}')\sum_\lambda A_\lambda(\mathbf{p},\mathbf{r}')B_\lambda(\mathbf{p},\mathbf{r}')\,d^3\mathbf{p}=0. \quad . \quad . \quad . \quad . \quad (10)$$

In generalizing these results for N particles, $\mathbf{r}$ is replaced by the set $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$. The effect of the symmetry restriction (Bose or Fermi) is that the ρ of (8) is in practice confined to the appropriate subspace of the space of $\{\phi_\nu\}$. The restriction is formally included by replacing all observables A by AQ, where $Q=Q^+$ or Q^- is the corresponding projection operator—which commutes with all A of physical interest. The correspondence (10) is not affected by the restriction.

§3. The Equations of Motion.

Write Schrödinger's equation in the form

$$H\psi(x_1, x_2, \ldots, x_{3N}) \equiv [-\omega\Delta+U(x_1, x_2, \ldots, x_{3N})]\psi=i\hbar\frac{\partial}{\partial t}\psi, \quad . \quad . \quad . \quad (11)$$

where $\omega=\hbar^2/2m$, m is the mass of a particle, and

$$\Delta\equiv\sum_{i=1}^{3N}\partial_i\partial_i, \qquad \partial_i\equiv\partial/\partial x_i.$$

Let

$$\mathbf{k}(x_1 \ldots x_{3N})\equiv\tfrac{1}{2}i\hbar[\psi\nabla\psi^*-\psi^*\nabla\psi]. \quad . \quad . \quad . \quad . \quad (12)$$

Then

$$\nabla\,.\,\mathbf{k}\equiv\sum_i\partial_i k_i=(\tfrac{1}{2}i\hbar/\omega)[\psi^*H\psi-\psi H\psi^*]=-m[\psi^*\dot{\psi}+\psi\dot{\psi}^*].$$

I. e.

$$\nabla\,.\,\mathbf{k}=-m\frac{\partial}{\partial t}(\psi^*\psi). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Also,

$$2\frac{\partial}{\partial t}\mathbf{k}=[(H\psi)\nabla\psi^*-\psi\nabla(H\psi^*)]-[(-H\psi^*)\nabla\psi+\psi^*\nabla(H\psi)]$$

$$=-2\psi^*\psi\nabla U-\omega[(\Delta\psi\nabla\psi^*+\Delta\psi^*\nabla\psi)-(\psi\nabla(\Delta\psi^*)+\psi^*\nabla(\Delta\psi))].$$

Hence, by the identity

$$\Delta\psi^*\nabla\psi+\Delta\psi\nabla\psi^*\equiv\nabla\,.\,[\nabla\psi^*\nabla\psi+\nabla\psi\nabla\psi^*]-\nabla(\nabla\psi^*\,.\,\nabla\psi),$$

$$\frac{\partial}{\partial t}\mathbf{k}+\psi^*\psi\nabla U=-\omega\{\nabla\,.\,[\nabla\psi^*\nabla\psi+\nabla\psi\nabla\psi^*]$$

$$-\tfrac{1}{2}\nabla[\psi^*\Delta\psi+\psi\Delta\psi^*+2\nabla\psi^*\,.\,\nabla\psi]\}$$

$$=-\omega\{\nabla\,.\,[\nabla\psi^*\nabla\psi+\nabla\psi\nabla\psi^*]-\tfrac{1}{2}\nabla[\Delta(\psi^*\psi)]\}.$$

I. e.

$$\frac{\partial}{\partial t}\mathbf{k}+\psi^*\psi\nabla U+\omega\nabla\,.\,[(\nabla\psi^*\nabla\psi+\nabla\psi\nabla\psi^*)-\tfrac{1}{2}\nabla\nabla(\psi^*\psi)]=0. \quad . \quad . \quad (14)$$

§ 4. THE MACROSCOPIC EQUATIONS.

Rewrite the wave-function in terms of the coordinates of the particles :

$$\psi=\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N),$$

where the $\mathbf{r}_i$ are vectors in subspaces of the cartesian space of § 3. If the surface S enclosing the particles is given by $\sigma(\mathbf{r})=0$, then this 3N-dimensional space is bounded by the hypersurfaces $S_i : \sigma(\mathbf{r}_i)=0$. The boundary condition is taken as

$$\psi=0 \text{ on the } S_i. \qquad \ldots \ldots \ldots \quad (15)$$

We may also define a " vector " element of S_i: $d\mathbf{S}_i$

$$\left(\text{so that, for example, } \int^{S_i} \mathbf{r}_i \,.\, d\mathbf{S}_i = 3V \prod_{j\neq i} d^3\mathbf{r}_j\right),$$

and vector operators ∇_i, $(\nabla_i \,.\, \mathbf{r}_j=3\delta_{ij})$. The hamiltonian of (11) is then

$$H=-\omega \sum_i \Delta_i + U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N). \qquad \ldots \ldots \quad (16)$$

Let I_i' denote the operation

$$\int^{\mathfrak{R}_1} \ldots \int^{\mathfrak{R}_N} \prod_{j\neq i} d^3\mathbf{r}_j \ldots,$$

and I_i mean applying I_i' and dropping the suffix i, so that

$$I_i f(\mathbf{r}_i)=V^{(N-1)} f(\mathbf{r}).$$

I_{ij}'', etc., are defined similarly. Then

$$\begin{aligned} I_i'(\nabla \,.\, \mathbf{k}) &= \sum_j I_i'(\nabla_j \,.\, \mathbf{k}_j) \\ &= I_i'(\nabla_i \,.\, \mathbf{k}_i) + \sum_{j\neq i} I_{ij}'' \int^{S_j} \mathbf{k}_j \,.\, d\mathbf{S}_j \\ &= I_i'(\nabla_i \,.\, \mathbf{k}_i) = \nabla_i \,.\, (I_i'\mathbf{k}_i), \text{ by (15).} \end{aligned}$$

Hence

$$I_i(\nabla \,.\, \mathbf{k})=\nabla \,.\, (I_i \mathbf{k}_i).$$

Defining

$$\mathbf{J} \equiv \sum_i I_i \mathbf{k}_i, \qquad n \equiv \sum_i I_i(\psi^*\psi), \qquad \ldots \ldots \quad (17)$$

(13) becomes

$$\nabla \,.\, \mathbf{J} = -m \frac{\partial}{\partial t} n. \qquad \ldots \ldots \ldots \quad (18)$$

Since, apart from normalizing constants, $\mathbf{J} \rightarrow mn_1\mathbf{u}_1^{(1)}$ (*cf.* Green's I, 5.1) and $n \rightarrow n_1$ (I, 2.7), (18) is equivalent to (I, 5.6) with $q=1$. Green's more general cluster equations correspond to integration of (13) over fewer than (N−1) of the coordinates.

Since (14) is a vector equation, we may write it

$$-I_i'\left[\frac{\partial}{\partial t}\mathbf{k}_i + \psi^*\psi \nabla_i U\right] = \sum_j I_i'(\nabla_j \,.\, \mathbf{T}_{ji}), \qquad \ldots \ldots \quad (19)$$

where

$$\mathbf{T}_{ji}=\nabla_j \psi^* \nabla_i \psi + \nabla_j \psi \nabla_i \psi^* - \tfrac{1}{2}\nabla_j \nabla_i(\psi^*\psi).$$

Now, for $j\neq i$,

$$\begin{aligned} I_i'(\nabla_j \,.\, \mathbf{T}_{ji}) = I_{ij}'' \int^{S_j} d\mathbf{S}_j \,.\, [(&\nabla_j \psi^* \nabla_i \psi + \nabla_j \psi \nabla_i \psi^*) \\ &- \tfrac{1}{2}\nabla_j(\psi^*\nabla_i\psi + \psi\nabla_i\psi^*)]=0, \text{ by (15),} \end{aligned}$$

since $\psi=0$, $\nabla_i\psi=0$, on S_j.

Defining

$$\left.\begin{aligned}
&\mathbf{R}\equiv \omega \sum_i \mathrm{I}_i[\nabla_i\psi^*\nabla_i\psi+\nabla_i\psi\nabla_i\psi^*],\\
&\mathbf{S}\equiv \tfrac{1}{2}\omega \sum_i \mathrm{I}_i\nabla_i\nabla_i(\psi^*\psi)=\tfrac{1}{2}\omega\nabla\nabla n,\\
&\mathbf{P}\equiv \mathbf{R}-\mathbf{S},\\
&\mathbf{F}\equiv -\sum_i \mathrm{I}_i(\psi^*\psi\nabla_i \mathrm{U}),
\end{aligned}\right\} \quad . \;\; . \;\; . \;\; . \;\; . \;\; . \qquad (20)$$

(19) becomes $$\frac{\partial}{\partial t}\mathbf{J}=\mathbf{F}-\nabla\,.\,\mathbf{P}. \quad . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \qquad (21)$$

This result is equivalent to (I, 5.8) for $q=1$, since

$$\mathbf{J}\rightarrow mn_1\mathbf{u}_1^{(1)}, \qquad \mathbf{F}\rightarrow mn_1\boldsymbol{\eta}_1^{(1)},$$

while for the remaining term of (I, 5.8) (there is a factor $(1/m^2)$ missing) :

$$(1/m)\{\{\rho_1, \mathbf{p}^{(1)}\}\mathbf{p}^{(1)}\}\rightarrow -\tfrac{1}{2}\omega\sum_i \mathrm{I}_i[(\psi^*\nabla_i\nabla_i\psi+\psi\nabla_i\nabla_i\psi^*)-(\nabla_i\psi^*\nabla_i\psi+\nabla_i\psi\nabla_i\psi^*)]$$

$$=\omega\sum_i \mathrm{I}_i[(\nabla_i\psi^*\nabla_i\psi+\nabla_i\psi\nabla_i\psi^*)-\tfrac{1}{2}\nabla_i\nabla_i(\psi^*\psi)]=\mathbf{P}.$$

The equation corresponding to (I, 5.10) follows from (21) in the same way, given the definitions of $\mathbf{v}$ and d/dt, and the equation of continuity (18). The quantity $\mathbf{P}$ defined in (20) is the expectation, in Landau's formalism, of

$$\left(\frac{1}{m}\right)\sum_i\{\{\mathbf{p}_i, \delta(\mathbf{r}', \mathbf{r}_i)\}, \mathbf{p}_i\},$$

which (by the considerations leading to (10)) reduces, in the classical limit, to the classical tensor of the momentum flux. Also $\mathbf{S}$ is proportional to the expectation of

$$\sum_i[[\mathbf{p}_i, \delta(\mathbf{r}', \mathbf{r}_i)], \mathbf{p}_i] \qquad (\text{where } [\mathrm{A}, \mathrm{B}]=(i\hbar)^{-1}(\mathrm{AB}-\mathrm{BA}))$$

and hence vanishes " as h^2 ", in the classical limit.

§ 5. THE VIRIAL THEOREM.

From the definitions (20), the scalar of $\mathbf{S}$ is

$$\mathbf{S}_s=\tfrac{1}{2}\omega\sum_i \mathrm{I}_i\Delta_i(\psi^*\psi)=\tfrac{1}{2}\omega\Delta n, \quad . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \qquad (22)$$

and similarly

$$\mathbf{R}_s=\omega\sum_i \mathrm{I}_i[\Delta_i(\psi^*\psi)-(\psi^*\Delta_i\psi+\psi\Delta_i\psi^*)]=\omega\Delta n+2\mathrm{K},$$

where

$$\mathrm{K}\equiv -\omega\sum_i \mathrm{I}_i(\psi^*\Delta_i\psi+\psi\Delta_i\psi^*), \quad . \;\; . \;\; . \;\; . \;\; . \;\; . \qquad (23)$$

is the kinetic energy expectation. Then from (21),

$$\nabla\,.\,(\mathbf{P}\,.\,\mathbf{r})=(\nabla\,.\,\mathbf{P})\,.\,\mathbf{r}+\mathbf{P}_s=\mathbf{F}\,.\,\mathbf{r}-\frac{\partial}{\partial t}\mathbf{J}\,.\,\mathbf{r}+2\mathrm{K}+\tfrac{1}{2}\omega\Delta n. \quad . \qquad (24)$$

For a stationary state, $(\partial/\partial t)\mathbf{J}=0$. Then, on integrating (24) over the whole volume,

$$\int^{s} d\mathbf{S} \,.\, \mathbf{P} \,.\, \mathbf{r}=\int^{V}(2\mathrm{K}-\Gamma)\,d^3\mathbf{r}, \qquad . \quad . \quad . \quad . \quad . \quad (25)$$

where $\Gamma\equiv-\mathbf{F}\,.\,\mathbf{r}$ is the virial of the potential U, and $\int^{V}(\Delta n)\,d^3\mathbf{r}$ vanishes by (15). (25) is the quantum virial theorem for an energy eigenstate, or a statistical ensemble of stationary states. Putting

$$\mathrm{U}=\mathrm{U}_1+\mathrm{U}_2\equiv\sum_i \mathrm{V}_1(\mathbf{r}_i)+\sum\sum_{i<j}\mathrm{V}_2(r_{ij}), \qquad \mathbf{r}_{ij}\equiv\mathbf{r}_i-\mathbf{r}_j, \qquad . \quad (26)$$

the virial becomes

$$\Gamma=\sum_i \mathrm{I}_i[\mathbf{r}_i\,.\,(\nabla_i \mathrm{V}_1(\mathbf{r}_i))\psi^*\psi]+\sum\sum_{i<j}\mathrm{I}_i[r_{ij}(\mathrm{V}_2'(r_{ij}))\psi^*\psi]. \quad . \quad . \quad (27)$$

Or

$$\Gamma=\mathbf{r}\,.\,(\nabla \mathrm{V}_1(\mathbf{r}))n(\mathbf{r})+\tfrac{1}{2}\int s\mathrm{V}_2'(s)n_2(\mathbf{r},\mathbf{r}+\mathbf{s})\,d^3\mathbf{s}, \quad . \quad . \quad . \quad (28)$$

where

$$n_2(\mathbf{r}_1,\mathbf{r}_2)=2\sum\sum_{i<j}\mathrm{I}_{ij}(\psi^*\psi).$$

By (15), on the boundary

$$\nabla_i\psi=\mathbf{n}_i\partial_{ni}\psi,$$

where $\mathbf{n}_i$ is the unit (outward) normal to S_i; and hence

$$\text{on S,} \quad \mathbf{P}=\mathbf{n}\mathbf{n}\omega\sum_i \mathrm{I}_i(\partial_{ni}\psi^*\partial_{ni}\psi). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

The interpretation of (29) as the tensor of the stress on the boundary S, suggested by the form of (25), is supported by the following considerations for one dimension.

The Schrödinger equation in one dimension is

$$\left[-\omega\frac{d^2}{dx^2}+\mathrm{U}(x)-\mathrm{E}\right]\psi(x)=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where

$$\lim_{x\to\pm\infty}\mathrm{U}>\mathrm{E}.$$

Suppose that, for $x>0$, the potential has the form

$$\left.\begin{array}{ll} x\leqslant a, & \mathrm{U}=0, \\ x>a, & \mathrm{U}=\mathrm{E}_0=\text{constant}, \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where $\mathrm{E}_0>\mathrm{E}$. Then the solution of (30) for $x>0$ is

$$\left.\begin{array}{ll} x\leqslant a, & \psi=\mathrm{A}\sin[k_1(x-a)+\phi], \\ x>a, & \psi=\mathrm{A}\sin(\phi)\exp[-k_2(x-a)], \\ \omega k_1^2=\mathrm{E}, & \omega k_2^2=\mathrm{E}_0-\mathrm{E}. \end{array}\right\} \quad . \quad . \quad (32)$$

Since $(d/dx)\psi$ is continuous,

$$\tan(\phi)=-k_1/k_2=-\sqrt{\left(\frac{\mathrm{E}}{\mathrm{E}_0-\mathrm{E}}\right)}. \quad . \quad . \quad . \quad . \quad . \quad (33)$$

The contribution of the "potential step" (31) to the reaction of the system is

$$\mathrm{F}=\lim_{\varepsilon\to 0}\int_{x=a-\varepsilon}^{a+\varepsilon}\psi^*\psi\,d\mathrm{U}=\mathrm{E}_0(\psi^*\psi)_{x=a}$$

$$=|\mathrm{A}|^2\mathrm{E}_0\sin^2(\phi)=\omega\left(\frac{d}{dx}\psi^*\right)_a\left(\frac{d}{dx}\psi\right)_a(\mathrm{E}_0\tan^2(\phi)/\mathrm{E}).$$

If E is fixed while $\mathrm{E}_0\to\infty$,

$$\left.\begin{aligned}&(a)\ \psi\to 0,\qquad x\geqslant a,\\&(b)\ \mathrm{E}_0\tan^2(\phi)/\mathrm{E}\to 1,\end{aligned}\right\}\quad . \quad . \quad . \quad (34)$$

and hence

$$\mathrm{F}\to\omega\left(\frac{d}{dx}\psi^*\right)_a\left(\frac{d}{dx}\psi\right)_a,\quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

which is the one-dimensional analogue of (29).

§ 6. APPLICATION OF FRÖHLICH'S PERTURBATION THEORY.

The interpretation of (29) suggested above, is confirmed by the theory of boundary perturbations proposed by Fröhlich (1938). In the notation of (11), let ψ satisfy $\mathrm{H}\psi=\mathrm{E}\psi$ in $\mathfrak{R}$ and vanish on its boundary S, and let ψ' satisfy $\mathrm{H}\psi'=\mathrm{E}'\psi'$ in $\mathfrak{R}'\subset\mathfrak{R}$ and vanish on S′. Then by Green's theorem

$$(\mathrm{E}'-\mathrm{E})\int^{\mathfrak{R}'}\psi^*\psi\prod_i dx_i=-\omega\int^{\mathrm{S}'}\psi^*\partial_n\psi'\,d\mathrm{S}.\quad . \quad . \quad . \quad (36)$$

Choose ψ' so that, as $\mathrm{S}'\to\mathrm{S}$, $\psi'\to\psi$ and $\mathrm{E}'\to\mathrm{E}$. Suppose that, if $\mathbf{x}$ is on S, $\mathbf{x}'=\mathbf{x}+\lambda\mathbf{z}(\mathbf{x})$ is on S′. Then it is assumed (Fröhlich 1938) that

$$\partial_n\psi'(\mathbf{x}+\lambda\mathbf{z})=\partial_n\psi(\mathbf{x})+\mathrm{O}(\lambda)\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

(*i. e.* ψ "moves with the boundary", to the first order in λ). Then (36) gives

$$\mathrm{E}'=\mathrm{E}-\lambda\omega\int^{\mathrm{S}}(\partial_n\psi^*)(\partial_n\psi)\mathbf{z}\,.\,d\mathrm{S}+\mathrm{O}(\lambda^2).\quad . \quad . \quad . \quad . \quad (38)$$

In the notation of § 4, if the surface S enclosing the particles is perturbed to S′ by $\mathbf{r}'=\mathbf{r}+\lambda\mathbf{s}(\mathbf{r})$, then (38) integrates to

$$\left(\frac{d\mathrm{E}}{d\lambda}\right)_{\lambda=0}=-\omega\int^{\mathrm{S}}\left[\sum_i\mathrm{I}_i(\partial_{ni}\psi^*)(\partial_{ni}\psi)\right]\mathbf{s}\,.\,d\mathrm{S}=-\int^{\mathrm{S}}d\mathrm{S}\,.\,\mathbf{P}\,.\,\mathbf{s}.\quad . \quad . \quad (39)$$

Conservation of energy then requires **P** to be the tensor of the stress on the boundary in a *slow* perturbation, provided Ehrenfest's theorem applies. For a Gibbs ensemble, the free energy F is given by

$$e^{-\beta\mathrm{F}}=\mathrm{Z}\equiv\sum_{\nu=1}^{\infty}e^{-\beta\mathrm{E}_\nu},\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

where E_ν are the energy levels of the eigen-functions of H, of Bose and Fermi symmetry, and $\beta^{-1}=kT$. Then

$$\left(\frac{\partial F}{\partial\lambda}\right)_\beta = Z^{-1}\sum_\nu e^{-\beta E_\nu}\frac{dE_\nu}{d\lambda} = -\int^S d\mathbf{S}.\left[Z^{-1}\sum_\nu e^{-\beta E_\nu}\mathbf{P}_\nu\right].\mathbf{s}, \qquad . \quad (41)$$

and therefore

$$\overline{\mathbf{P}} \equiv Z^{-1}\sum_\nu e^{-\beta E_\nu}\mathbf{P}_\nu,$$

the average of $\mathbf{P}$ over the canonical distribution, is the boundary stress for isothermal displacements. For a simple dilation

$$\mathbf{s}=\mathbf{r}, \quad V\frac{d}{dV}=\left(\frac{1}{3}\right)\frac{d}{d\lambda}. \qquad . \; . \; . \; . \; . \; . \; . \quad (42)$$

Substituting (42) in (25), (39) and (41), the thermodynamic pressure is

$$p=-\frac{dF}{dV}=\left(\frac{1}{3V}\right)\int^V\left[Z^{-1}\sum_\nu e^{-\beta E_\nu}(2K_\nu-\Gamma_\nu)\right]d^3\mathbf{r}.$$

I. e.

$$3pV=\int^V(2\overline{K}-\overline{\Gamma})d^3\mathbf{r}. \qquad . \; . \; . \; . \; . \; . \; . \; . \quad (43)$$

There is an arbitrariness in (39), which can be written

$$-\frac{dE}{d\lambda}=\int_S d\mathbf{S}.(\alpha\mathbf{R}-\beta\mathbf{S}).\mathbf{s}$$

where $2\alpha-\beta=1$. For the case (42), (43), $\mathbf{s}=\mathbf{r}$ we then find

$$-\frac{dE}{d\lambda}=\int^V(2K-\Gamma)d^3\mathbf{r}+3\left(\frac{\alpha-1}{\alpha}\right)\int^V \mathbf{S}_s d^3\mathbf{r}, \quad \alpha\neq 0.$$

Since $\int^V \mathbf{S}_s d^3\mathbf{r}=0$ by (15), we then retrieve (43).

Comparison of (23), (28) and (43) with (I, 3.21) and (I, 3.22) shows that p is numerically equal to Green's p_1, the first term in his series for p. Thus the quantity $\pi=p_1-p$, which Green identifies (II) with the energy associated with the thermomechanical effect in liquid helium II, appears to be zero. However, this conclusion has been the subject of a recent controversy (Yvon 1948), (Massignon 1949), (de Boer 1949), (Green 1949), (Horowitz 1950). A further discussion of the question is therefore attempted in the following paragraphs.

§ 7. APPLICATION OF BRILLOUIN'S THEORY.

The result (39) may also be obtained by the independent perturbation method originated by Brillouin (1937). For simplicity, the work is given in terms of the case N=1. The particle is enclosed in the surface S, on which the wave-function $\psi(\mathbf{r})$ vanishes. Let S be continuously deformed to S′ by $\mathbf{r}'=\mathbf{r}+\lambda\mathbf{s}(\mathbf{r})$, and choose $\mathbf{s}$ to be normal to S. Then $\mathbf{s}(\mathbf{r})$ may be chosen *inside* S so as to agree with the boundary value, be continuous, and satisfy

$$\nabla.\mathbf{s}=\text{constant}, \quad \mathbf{s}=\nabla\phi. \qquad . \; . \; . \; . \; . \; . \; . \quad (44)$$

The choice is unique, and has the property

$$\Delta\mathbf{s}=0. \qquad (45)$$

If $$H'\psi(\mathbf{r})=H\psi(\mathbf{r}')=H\psi(\mathbf{r}+\lambda\mathbf{s}),$$

then $$H'=H+\lambda\{\omega[2(\nabla\mathbf{s}):\nabla\nabla+(\Delta\mathbf{s}).\nabla]+\mathbf{s}.\nabla U\}+0(\lambda^2).$$

Or, by (45),

$$G\equiv\left(\frac{dH}{d\lambda}\right)_{\lambda=0}=2\omega(\nabla\mathbf{s}):\nabla\nabla+[\mathbf{s}.\nabla U(\mathbf{r})]. \qquad (46)$$

Since, by (44), the eigenfunctions, ψ_ν, of H remain orthogonal under the deformation, the change in the eigenvalues is given by

$$\frac{dE_\nu}{d\lambda}=(\psi_\nu|G|\psi_\nu). \qquad (47)$$

Now

$$\begin{aligned}\nabla.(\mathbf{P}.\mathbf{s})&=(\nabla.\mathbf{P}).\mathbf{s}+\mathbf{P}:\nabla\mathbf{s}\\&=-\psi^*\psi(\nabla U).\mathbf{s}+\omega\nabla\mathbf{s}:[\nabla\psi^*\nabla\psi+\nabla\psi\nabla\psi^*-\tfrac{1}{2}\nabla\nabla(\psi^*\psi)]\end{aligned}$$

for a stationary state. Then, integrating by parts and using (45) and the boundary condition, we find

$$\begin{aligned}-\int^{S}d\mathbf{S}.\mathbf{P}.\mathbf{s}&=\int^{V}[\psi^*\psi(\nabla U).\mathbf{s}]d^3\mathbf{r}+2\omega\int^{V}[\nabla\mathbf{s}:\psi^*\nabla\nabla\psi]d^3\mathbf{r}\\&=(\psi|G|\psi)=\left(\psi\left|\frac{dH}{d\lambda}\right|\psi\right). \qquad (48)\end{aligned}$$

The generalization for N particles is quite straightforward. Thus (39) is confirmed, in a form which covers degenerate unperturbed eigenstates. If the boundary wall exerts a short-range force on the molecules, with potential U_3 which " moves with the wall " in the perturbation, then the derivation of (46) shows that the boundary stress is still given by (46) and (47) if $\mathbf{s}.\nabla U$ is replaced by

$$\mathbf{s}.\nabla(U_1+U_2)+z\left(\frac{dU_3(z)}{dz}\right)\partial_n s_n, \qquad (49)$$

where $\mathbf{n}$ is the unit normal to S, and $-z\mathbf{n}$ is the vector from $\mathbf{r}$ to S. For the dilation case, $\mathbf{s}=\mathbf{r}$, (49) reduces to $\mathbf{r}.\nabla U$, and (43) is still valid.

§ 8. THE THERMODYNAMIC-PRESSURE SERIES.

Derivations of (43) have been given (de Boer 1949), (Green 1949) which make use of the transformation of the partition sum Z of (40) to

$$Z=\mathrm{Lim}(p,n)\sum_{k=1}^{p}\langle k|\sigma_n|k\rangle \qquad (50)$$

where $$\sigma_n=\sum_{r=0}^{n}\frac{(-\beta)^r}{r!}H^r, \qquad (51)$$

$\{|k\rangle\}$ is any complete (in the Bose or Fermi subspace) normalized orthogonal basis, and $\mathrm{Lim}(p,n)$ means

$$\lim_{p\to\infty}\quad\lim_{n\to\infty}$$

taken in the order from right to left (it is, of course, not permissible to interchange the limits in (50)). By (46),

$$\frac{d}{d\lambda}(\mathrm{H}^{r+1})=\mathrm{H}^r\mathrm{G}+\mathrm{H}^{r-1}\mathrm{GH}+\ .\ .\ .\ +\mathrm{GH}^r=\sum_{t=0}^{r}\binom{r+1}{t+1}\mathrm{G}_t\mathrm{H}^{(r-t)}. \quad (52)$$

where

$$\mathrm{G}_t=\mathrm{HG}_{t-1}-\mathrm{G}_{t-1}\mathrm{H},\quad \mathrm{G}_0=\mathrm{G}.$$

Hence, by (50),

$$\frac{d\Xi}{d\lambda}=\mathrm{Lim}(p,n)\sum_{k=1}^{p}\sum_{r=1}^{n}\sum_{t=1}^{r}\left[\frac{(-\beta)^r}{t!(r-t)!}\langle k\,|\,\mathrm{G}_{t-1}\mathrm{H}^{(r-t)}\,|\,k\rangle\right].\quad .\quad (53)$$

If we choose, for the $|\,k\rangle$, the eigenkets of the *unperturbed* hamiltonian, then

$$\langle k\,|\,\mathrm{G}_t\mathrm{H}^r\,|\,k\rangle=\langle k\,|\,\mathrm{HG}_{t-1}\,|\,k\rangle\mathrm{E}_k{}^r-\langle k\,|\,\mathrm{G}_{t-1}\,|\,k\rangle\mathrm{E}_k{}^{(r+1)}.$$

If, H being hermitian, we put

$$\langle k\,|\,\mathrm{HG}_{t-1}\,|\,k\rangle=\mathrm{E}_k\langle k\,|\,\mathrm{G}_{t-1}\,|\,k\rangle$$

(it is pointed out below that this step requires a special justification), then

$$\langle k\,|\,\mathrm{G}_t\mathrm{H}^r\,|\,k\rangle=0,\quad t\geqslant 1.\quad .\ .\ .\ .\ .\ .\ . \quad (54)$$

Substituting (54) in (53), by (48) we retrieve (41) and (43).

Since (for $\mathbf{s}=\mathbf{r}$)

$$\mathrm{G}=3\mathrm{V}\frac{d\mathrm{H}}{d\mathrm{V}}=(i\hbar)^{-1}(\mathrm{H}\mathbf{p}\,.\,\mathbf{r}-\mathbf{p}\,.\,\mathbf{r}\mathrm{H}),\quad .\ .\ .\ . \quad (55)$$

the argument leading to (54) also requires $\langle k\,|\,\mathrm{GH}^r\,|\,k\rangle$ to vanish. It has been pointed out, in this connection, by Green (1949) that expressions of the form

$$\mathrm{Tr}(\mathrm{AB}-\mathrm{BA})=\mathrm{Lim}(n,\nu)\sum_{r=1}^{n}\sum_{\varrho=1}^{\nu}[\mathrm{A}_{r\varrho}\mathrm{B}_{\varrho r}-\mathrm{B}_{r\varrho}\mathrm{A}_{\varrho r}]$$

need not vanish, with a dense or continuous basis $\{|\,r\rangle\}$, if $\langle r\,|\,\mathrm{AB}\,|\,r\rangle$ is not uniformly convergent with respect to r. There is, however, another approach to such questions, which was pointed out by Fuchs (1940). H is specified not simply by a differential operator, such as (11), but with respect to the linear space of a basis $\{|\,k\rangle\}$—*i. e.* to the domain symmetry and boundary condition of the corresponding wave-functions. H is then only hermitian in virtue of this boundary condition, and if $\mathrm{A}\,|\,k\rangle$ does not satisfy the latter (A is then called by Fuchs " inadmissible ") we cannot assume that

$$\langle j\,|\,(\mathrm{HA}\,|\,k\rangle)=[\langle k\,|\,(\mathrm{AH}\,|\,j\rangle)]^*.$$

Differential relations such as (46) and (55) are still correct, but deductions by matrix algebra may be invalid. For example, a virial theorem may be deduced as follows :

$$m\frac{d^2}{dt^2}(r_i{}^2)=m\frac{d}{dt}[r_i{}^2,\mathrm{H}]=\frac{d}{dt}(\mathbf{r}_i\,.\,\mathbf{p}_i+\mathbf{p}_i\,.\,\mathbf{r}_i)$$
$$=\ .\ \ .\ \ .\ \ =2[p_i{}^2/m-\mathbf{r}_i\,.\,\nabla_i\mathrm{U}],$$

and hence for a stationary state

$$\langle | \Sigma i(p_i^2/m - \mathbf{r}_i \,.\, \nabla_i \mathrm{U}) | \rangle = 0.$$

The surface-integral term is missing because, although the commutator relation $[\mathbf{r}_i, \mathbf{p}_j] = \mathsf{l}_i \delta_{ij}$ is applicable to $\mathbf{r}_i$ and $\mathbf{p}_j$ *as differential operators*, the Heisenberg equation

$$\frac{d}{dt} \langle | \mathrm{A} | \rangle = \langle | [\mathrm{A}, \mathrm{H}] | \rangle$$

is then invalid where A is not admissible. Now, with the boundary condition $\psi = 0$ which we are considering, G and the G_t are admissible, since G—2H reduces on the boundary to a function of $\mathbf{x}$, but $\mathbf{p} \,.\, \mathbf{x} + \mathbf{x} \,.\, \mathbf{p}$ is not. It follows that the diagonal elements, in the energy representation, of G_1, G_2, . . . necessarily vanish, but not those of G.

§ 9. THE COORDINATE REPRESENTATION.

Green's (1949) treatment is based on the Schrödinger representation of (50) (Slater 1931, 1933):

$$Z = \int e^{-\beta \mathrm{H}}(\mathbf{x}', \mathbf{x}') d^{3\mathrm{N}}\mathbf{x}'. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (56)$$

A typical term in the resulting series for $d\mathrm{F}/d\lambda$ (simplified by not employing the relation (55), and not symmetrizing the operators in each term) is proportional to

$$\int d^{3\mathrm{N}}\mathbf{x}' [\mathrm{G}_t e^{-\beta \mathrm{H}}](\mathbf{x}', \mathbf{x}') = \int d^{3\mathrm{N}}\mathbf{x}' \sum_{\nu=1}^{\infty} \{\psi_\nu^*(\mathbf{x}') [\mathrm{G}_t \psi_\nu(\mathbf{x}')] e^{-\beta \mathrm{E}_\nu}\}, \quad . \qquad (57)$$

where the ψ_ν are the eigenfunctions of H. For $t \geqslant 1$, since $|\psi_\nu|$ is bounded and (at least in the dilation case (42)) G_t reduces to a finite function of position, (57) is

$$\sum_{\nu=1}^{\infty} e^{-\beta \mathrm{E}_\nu} \left[\int d^{3\mathrm{N}}\mathbf{x}' \psi_\nu^*(\mathbf{x}') [\mathrm{G}_t \psi_\nu(\mathbf{x}')] \right] = \sum_{\nu=1}^{\infty} e^{-\beta \mathrm{E}_\nu} \langle \nu | \mathrm{G}_t | \nu \rangle = 0.$$

Thus de Boer's (1949) statement, that Green's series reduces to its first term, is confirmed. A recent paper by Horowitz (1950) appears to suggest:

(*a*) that G is inadmissible, in Fuchs' sense,

(*b*) that the reasoning, corresponding to (53), which leads to Green's series is therefore invalid.

We have seen that G (though not Green's operator) *is* admissible for the boundary condition $\psi = 0$, which Horowitz also adopts in the context of his criticism (G is not admissible for a periodic boundary condition, which leads to difficulties if applied in this problem). But in any case, from the present point of view, the admissibility of G affects, not the validity of the derivation of Green's series, but the vanishing or non-vanishing of the terms of this series. However, the validity of the use of (50), in the form (56), needs examination—since the transformation matrix

$$\mathrm{T}_{\mathbf{x}'\nu} = \langle \mathbf{x}' | \nu \rangle = \psi_\nu(\mathbf{x}') \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (58)$$

is singular ($T^\dagger T(\mathbf{x}', \mathbf{x}'')=\delta(\mathbf{x}', \mathbf{x}'')$).

If
$$I(\mathfrak{S}, n, \lambda)=\int^{\mathfrak{S}} d^{3N}\mathbf{x}' \sum_{r=0}^{n} \sum_{\nu=1}^{\lambda} \frac{(-\beta)^r}{r\,!}\, \psi_\nu^*(\mathbf{x}')[H^r\psi_\nu(\mathbf{x}')]$$
$$=\int^{\mathfrak{S}} d^{3N}\mathbf{x}' \sum_{r=0}^{n} \sum_{\nu=1}^{\lambda} \frac{(-\beta)^r}{r\,!} E_\nu{}^r\psi_\nu^*(\mathbf{x}')\psi_\nu(\mathbf{x}'), \quad . \quad . \quad . \quad (59)$$

and Lim ($\mathfrak{S}$) means the limit $\mathfrak{S}\rightarrow\mathfrak{R}$, then of the six expressions

$$\left.\begin{aligned} Z_1&=\text{Lim}\,(n, \mathfrak{S}, \lambda)I,\\ Z_2&=\text{Lim}\,(n, \lambda, \mathfrak{S})I,\\ Z_3&=\text{Lim}\,(\lambda, n, \mathfrak{S})I,\\ Z_4&=\text{Lim}\,(\lambda, \mathfrak{S}, n)I,\\ Z_5&=\text{Lim}\,(\mathfrak{S}, \lambda, n)I,\\ Z_6&=\text{Lim}\,(\mathfrak{S}, n, \lambda)I, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

Z_1, Z_2 and Z_6 are divergent. Using the fact that the $|\psi_\nu|$ are bounded, $Z_3=Z_4=Z_5$. But
$$Z_3=\sum_{\nu=1}^{\infty} e^{-\beta E_\nu}=Z, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

while Z_5 is evidently the Slater (1933) sum, (56). Then, by (52) and (59),
$$\frac{dZ}{d\lambda}=\frac{dZ_5}{d\lambda}$$
$$=\text{Lim}\,(\mathfrak{S}, \lambda, n)\int^{\mathfrak{S}} d^{3N}\mathbf{x}' \sum_{r=1}^{n} \sum_{\nu=1}^{\lambda} \sum_{t=1}^{r} \frac{(-\beta)^r}{t\,!\,(r-t)\,!}\, \psi_\nu^*(\mathbf{x}')[G_{t-1}H^{(r-t)}\psi_\nu(\mathbf{x}')]$$
$$=\text{Lim}\,(\mathfrak{S}, \lambda, n)\int^{\mathfrak{S}} d^{3N}\mathbf{x}' \sum_{\nu=1}^{\lambda}\left\{ \sum_{r=1}^{n} \sum_{t=1}^{r} \frac{(-\beta)^r E_\nu{}^r}{t\,!(r-t)\,!}\, \psi_\nu^*(\mathbf{x}')[G_{t-1}\psi_\nu(\mathbf{x}')]\right\}. \quad . \quad . \quad . \quad (62)$$

Since
$$\sum_{s=0}^{\infty} \frac{(-\beta)^s}{s\,!} E_\nu{}^s\psi_\nu^*(\mathbf{x}')[G_{t-1}\psi_\nu(\mathbf{x}')]$$

converges uniformly with respect to t, $\mathbf{x}'$ and ν,
$$\text{Lim}\,(n)\{\ \}=\sum_{t=1}^{\infty} \frac{(-\beta)^t}{(t+1)\,!} e^{-\beta E_\nu}\psi_\nu^*(\mathbf{x}')[G_{t-1}\psi_\nu(\mathbf{x}')] \quad . \quad . \quad . \quad (63)$$

which also converges uniformly with respect to $\mathbf{x}'$ and ν, and hence
$$\frac{dZ}{d\lambda}=\text{Lim}(\mathfrak{S}, r)\int^{\mathfrak{S}} d^{3N}\mathbf{x}'\left[\sum_{t=1}^{r} \frac{(-\beta)^t}{t!} (G_{t-1}e^{-\beta H})\,(\mathbf{x}', \mathbf{x}')\right]. \quad . \quad . \quad (64)$$

Interchange of the remaining two limits results in
$$\frac{dZ}{d\lambda}=(-\beta)\sum_{t=0}^{\infty} \frac{(-\beta)^t}{t+1}\int^{\mathfrak{R}} d^{3N}\mathbf{x}'(G_t e^{-\beta H})(\mathbf{x}', \mathbf{x}'), \quad . \quad . \quad (65)$$

which is essentially Green's series, and which we have shown to reduce to its first term.

CONCLUSION.

Of the mathematical results in (I), the work in this paper confirms the dynamical equations (5.6), (5.11), and thus vindicates the formalism which led to them, but is at variance with the more controversial conclusion of §3: that the thermodynamic stress on the boundary (or on a "probe" inserted in the liquid) differs from the kinetic stress. However, it is not clear how far the theory in (II) really depends on (I). The formulæ for the thermodynamic stress necessarily refer to the whole volume of the system, the integrals over which involve a boundary contribution associated with $\mathbf{S}$, (20). The theory given in (II) of the thermo-mechanical effects (*cf.* II, 3.20) must be regarded as an empirical one (π was in fact computed from Kapitza's experimental results), while the discussion of the normal viscosity (I, 6.1) (II, §2) seems to depend on a conception, of the local order in the liquid, whose foundation is not clear. Further progress will probably depend on the development of a theory of the structure of the quantum liquid when almost in equilibrium. It may be, for example, that there is a long-range order which makes it meaningless to speak of local quasi-isothermal processes, to which results similar to (41) would apply.

I wish to express my thanks to Mr. H. N. V. Temperley, of the Royal Society Mond Laboratory, for his advice and encouragement, and I am indebted to the D.S.I.R. for a maintenance award.

REFERENCES.

BORN, M., and GREEN, H. S., 1947, *Proc. Roy. Soc.* A, **191**, 168.
BRILLOUIN, L., 1937, *Comptes Rendus*, **204**, 1863.
DE BOER, J., 1949, *Physica*, **15**, 843.
DIRAC, P. A. M., 1947, *Principles of Quantum Mechanics*, 3rd ed., p. 130 (Oxford: University Press).
FRÖHLICH, H., 1938, *Phys. Rev.*, **54**, 945.
FUCHS, K. E., 1940, *Proc. Roy. Soc.* A, **176**, 214.
GREEN, H. S., 1948, *Proc. Roy. Soc.* A, **194**, 244; 1949, *Physica*, **15**, 882.
HOROWITZ, J., 1950, *J. de Phys.*, **11**, 241.
LANDAU, L., 1941, *J. Phys. U.S.S.R.*, **5**, 71:
MASSIGNON, D., 1949, *Comptes Rendus*, **228**, 62.
SLATER, J., 1931, *Phys. Rev.*, **38**, 237; 1933, *J. Chem. Phys.*, **1**, 687.
VON NEUMANN, J., 1943, *Mathematische Grundlagen der Quanten Mechanik*, ch. 4. New York: Dover.
YVON, J., 1948, *Comptes Rendus*, **227**, 763.

LXXXVIII. CORRESPONDENCE.

Latitude Effect of Cosmic Ray Stars.

By S. Lattimore,
Imperial College of Science and Technology, London*.

[Received July 25, 1950.]

Ilford 200 μ G.5 emulsions were exposed vertically for 27 days at altitudes of 11,550 ft. and 41,400 ft. at a latitude of approximately 12° S., in the Bolivian Andes. The plates were exposed under a corrugated iron roof. Since the time required for transporting the plates to and from Bolivia was considerable, control plates of a similar type to those actually exposed were used to find the number of stars collected before and after the exposures.

The results are given in Table I., together with figures for the Jungfraujoch. The latter were calculated from results at 11,500 ft., obtained using 200 μ G.5 plates, assuming a range of 145±10 gm./cm.^{2} of air for the star producing radiation.

Table I.

Height (feet)	No. of stars (< 3 prong)/c.c./day.		Ratio Jungfrau/Andes
	Andes	Jungfrau	
11,550	7·7±1·5	16·4±0·5	2·1 ±0·4
14,400	13·4±1·7	27·5±1·3	2·05±0·27

The value of 2·1±0·4 for the ratio of the number of stars with three or more prongs at the Jungfraujoch to the number at 12° S. is in fair agreement with results of Beets, Morand and Winard (1949) who found a value of 2·4 for the ratio between 44° N. and 1° S. at a height of 14,700 ft.

In Table II. are shown the size distributions for all stars at the Andes, and those at the Jungfraujoch, normalized so that the ratio of all stars with three or more prongs at the Jungfraujoch to all stars with three or more prongs at the Andes is equal to 2·1. As can be seen, the latitude effect decreases as the star size increases.

Table II.

Star size	No. of Stars.		Ratio $\frac{\text{Jungfrau}}{\text{Andes}}=R$
	Andes	Jungfraujoch	
3, 4	193 ±25	498±22	2·6 ±0·35
5, 6	83·5±14	161±13	1·93±0·35
7, 8	42 ±9	52± 7	1·24±0·32
>8	41·5± 8·5	46± 7	1·10±0·27
Total	360	755	2·10

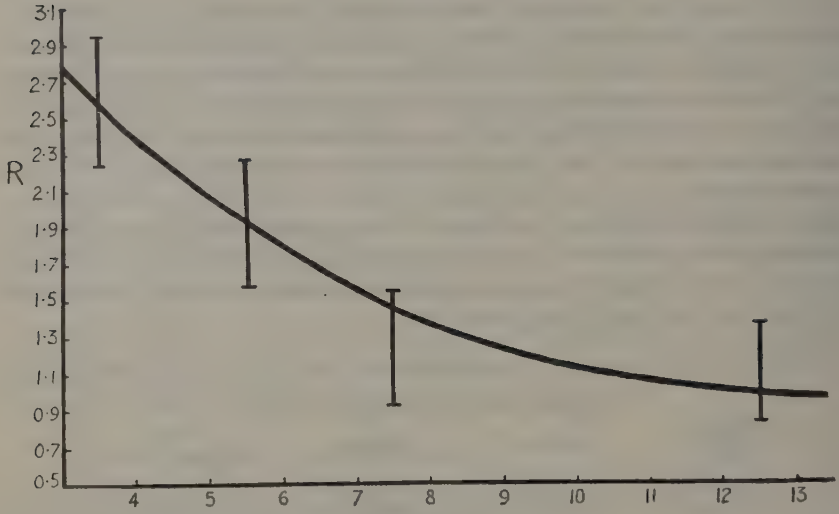
---

* Communicated by Professor Sir George Thomson, F.R.S.

The figures in Table II. are plotted graphically in fig. 1 and are shown to lie on a smooth curve.

Now the magnetic cut-off energy at the Andes is 15,000 MeV. and at the Jungfraujoch is 2000 MeV., and it is possible to estimate from the ratio of the number of stars at the Jungfraujoch to the number at the Andes the proportion of the stars at the Jungfraujoch that are produced by primaries with energies between 2000 and 15,000 MeV.

Fig. 1.



Star size.

Let x be this fraction. Then

$$2{\cdot}1 = 1/(1-x) \qquad \therefore \quad x \sim \tfrac{1}{2};$$

i.e. only one-half of the stars at the Jungfraujoch are produced by primaries with energies between 2000 and 15,000 MeV., although, assuming an E^{-3} energy spectrum for the primaries, there are approximately 50 times as many primaries in this range as there are with energy greater than 15,000 MeV. This result is, at first sight, rather surprising, since most of the stars have energies below 300 MeV. However, the majority of the stars are produced by neutral particles, probably neutrons, which must have been produced by the primaries high in the atmosphere, and further, the work of Harding (1949) shows that a star producing particle can probably produce more than one star. It would, therefore, seem probable that a primary produces several particles high in the atmosphere each of which can produce several stars, in which case the lower energy primaries would become less important.

ACKNOWLEDGMENTS.

We have pleasure in expressing our gratitude to Mr. Clifford Waite of Consolidated Tin Smelters Ltd., and to Mr. D. C. Deringer, Jr. of Platino Mines and Enterprises Cons. Inc. for exposing the Andes plates used in this experiment. Our thanks are also due to Sir Roger Makins, K.C.B., of the Foreign Office, who arranged for the transport of the plates to and from Bolivia. We are grateful to Professor Sir George Thomson for interesting discussions on this work.

Imperial College London. S. LATTIMORE.

REFERENCES.

BEETS, C., MORAND, M., and WINARD, L., 1949, *C.R. Acad. Sci.*, Paris, 229, 1227–8.

HARDING, J. B., 1949, *Phil. Mag.* (7) **40**, 530.

Note on the Heavy Nuclei of the Cosmic Radiation.

By A. D. DAINTON and D. W. KENT,
H. H. Wills Physical Laboratory, University of Bristol *.

[Received July 30, 1950.]

[Plates XXX. and XXXI.]

It is now well known that the atomic nuclei among the primary cosmic radiation are rapidly absorbed in the atmosphere as a result of nuclear interactions and loss of energy through ionization (Bradt and Peters 1948, 1950, Freier *et al.* 1948). Further, in the nuclear collisions which they make, these primary particles are often not absorbed catastrophically. They frequently suffer fragmentation so that fast heavy particles of smaller mass emerge from the encounter—see Pls. XXX. and XXXI. It follows that the distribution in mass and energy of the heavy particles, as well as their intensity, changes rapidly as they penetrate towards the earth.

Recent experiments in this laboratory with plates exposed at 68,000 feet, showed the presence of substantial numbers of heavy particles at this altitude. The observations have allowed us to determine the characteristics of the flux of heavy particles at this altitude. The deduced values of the intensity of the heavy nuclei at the top of the atmosphere, and their mean free path, are in good agreement with the results of other workers.

The exposure was made with 400 μ Ilford G5 emulsions by means of free balloons at a geomagnetic latitude of 54° N. Level flight was maintained at 68,000 feet (52 ± 1 gm./cm.^{2}) for 100 minutes. Since very few heavy nuclei are observed at altitudes lower than 55,000 feet, and since

---

[*] Communicated by Professor C. F. Powell, F.R.S.

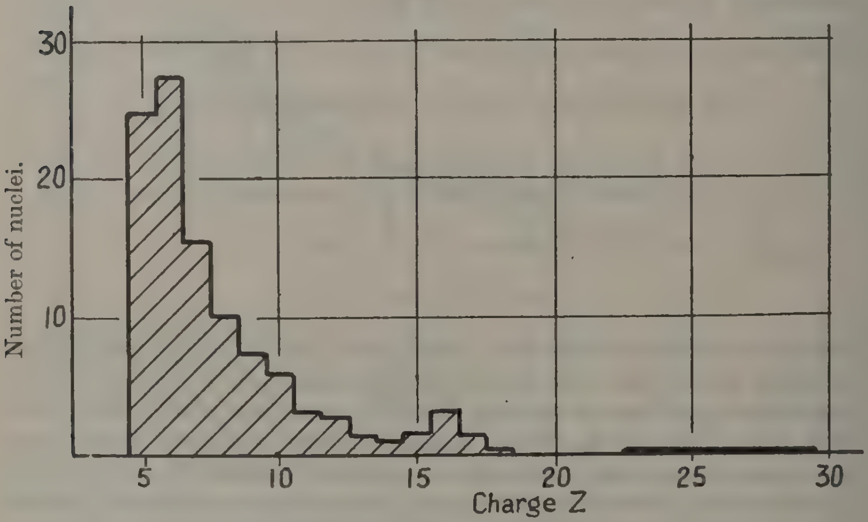
the time of ascent and descent above this altitude was small compared with the time of level flight, it was a sufficient approximation to assume that the observed nuclei entered the plates at 68,000 feet.

Our observations were made on 118 nuclei with charge $Z \geqslant 5$. The charge of each nucleus was determined by the delta-ray method described by Bradt and Peters (1948). The number, ν, of delta-rays having an energy greater than W_1, produced per unit length of its trajectory by a particle of charge Z, is given by :

$$\nu = \frac{CZ^2}{\beta^2}\left\{\frac{1}{W_1} - \frac{1}{\beta^2}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where βc is the velocity of the incident particle. Using the range-energy relation, the charge can thus be determined within an error of plus or minus one unit in the majority of cases, if the length of the track is sufficiently great. In cases where particles of low charge cannot be observed over a long range, the uncertainty in the charge determination is greater. The spectrum determined in this investigation is shown in fig. 1.

Fig. 1.



Distribution in charge of 118 nuclei.

In order to obtain a value of the flux at the top of the atmosphere it was convenient to consider only those tracks which passed through the emulsion of the outside plates of the stack. The angle of each track with respect to the zenith was determined from the dip and the angle between the projection of the track on the plane of the emulsion and the vertical side of the plate. The observed distribution was then corrected for absorption in the stack and for geometrical orientation of the plates. The number of tracks observed having a zenith angle between θ and $\theta + d\theta$ is given by

$$dn = I(\theta) \,.\, t \,.\, A \,.\, 4 \sin^2 \theta \,.\, f(\theta) \,.\, d\theta, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

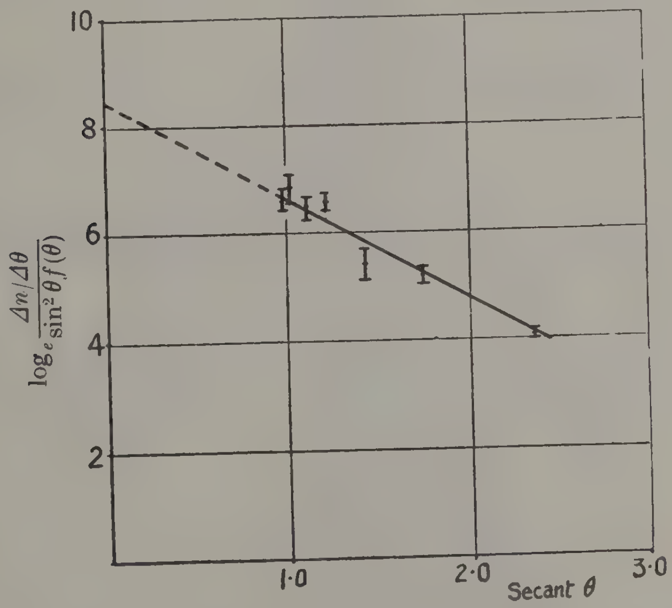
where t is the time of exposure, A is the area surveyed, and $4 \sin^2 \theta \,.\, f(\theta)$ is the correction factor due to plate orientation and to the choice of minimum track length (500 μ in this investigation). The function

$$f(\theta)=1-\left[1-\left(\frac{d}{l \sin \theta}\right)^2\right]^{1/2},$$

where d is the depth of the emulsion and l is the minimum track length considered, is operative only when the zenith angle exceeds $\sin^{-1} d/l=53°$. Assuming the absorption in the atmosphere to be exponential, $I(\theta)$, the intensity of the radiation at an angle θ, is related to the flux at the top of the atmosphere by the equation

$$I(\theta)=\mathrm{I}_0 \exp\left[(\mathrm{L}/\lambda) \sec \theta\right],$$

Fig. 2.



Angular distribution function.

L being the distance in gm./cm.^{2} from the level of the exposure to the top of the atmosphere, and λ the mean free path for disappearance in air. We can express equation (2) in the form (Hoang Tchang-Fong 1950)

$$\log_e\left[\frac{\Delta n}{\Delta \theta}\Big/\sin^2 \theta f(\theta)\right]=\log_e (4 \,.\, \mathrm{I}_0 \,.\, \mathrm{A} \,.\, t)-\frac{\mathrm{L}}{\lambda} \sec \theta.$$

A plot of this function, shown in fig. 2, gives a value $-\mathrm{L}/\lambda$ for the slope and for the intercept at $\sec \theta=0$, $\log_e (4 \,.\, \mathrm{I}_0 \,.\, \mathrm{A} \,.\, t)$. From these two values we obtain for the mean free path

$$\lambda \sim 28 \text{ gm./cm.}^2,$$

and for the flux $\mathrm{I}_0 \sim 1{\cdot}3 \times 10^{-3}$ particles cm.$^{-2}$ sec.$^{-1}$ steradian^{-1}.

It is probable that an appreciable fraction of the nuclei observed at this altitude are not of primary origin. Of eight observed collisions caused by nuclei of charge $Z > 6$, four cases resulted in a reduction in the charge of the primary, but little change in its velocity and direction.

In 34 cases in which an increase in the delta-ray density with range was observed, it was possible to determine the kinetic energy of the nucleus. The values thus obtained were corrected for ionization losses in the air to determine the corresponding values of the kinetic energy at the top of the atmosphere. Since collisions involving degradation of the charge occur, it is reasonable to assume that some of the particles are splinters of larger, more heavily ionizing nuclei. The values obtained, which lie between 0·45 BeV. and 2·0 BeV. per nucleon, therefore represent lower limits for the energies of the particles at the top of the atmosphere.

The present experiments are being continued with the application of new techniques.

The authors wish to express their gratitude to Professor C. F. Powell, F.R.S., for making this investigation possible, and to Dr. C. Franzinetti for helpful guidance and discussion.

REFERENCES.

BRADT, H. L., and PETERS, B., 1948, *Phys. Rev.*, **74,** 1828 ; 1950, *Ibid.*, **77,** 54.
FREIER, P., LOFGREN, E. J., NEY, E. P., and OPPENHEIMER, F., 1948, *Phys. Rev.*, **74,** 1818.
HOANG TCHANG-FONG, 1950, *Thesis*, Paris.

Loss of Energy of Fast Protons in Matter with Particular Reference to their Range in Carbon.

By A. E. TAYLOR,
Atomic Energy Research Establishment, Harwell, Berks.*

[Received August 3, 1950.]

IN using the 110-inch cyclotron at Harwell for measurements of total scattering cross sections and neutron energy spectra (Taylor, Pickavance, Cassels and Randle 1950) carbon absorbers of varying thickness have been used in a triple coincidence telescope detecting protons scattered from a polythene radiator. The range in carbon was used to deduce the energy of the protons, since carbon was a convenient material having a low atomic number. This choice was made in order to minimize multiple coulomb scattering and nuclear absorption for efficient operation of the detecting telescope. It was therefore required to compute the range energy relation in carbon for protons up to 200 MeV.

---

* Communicated by the Author.

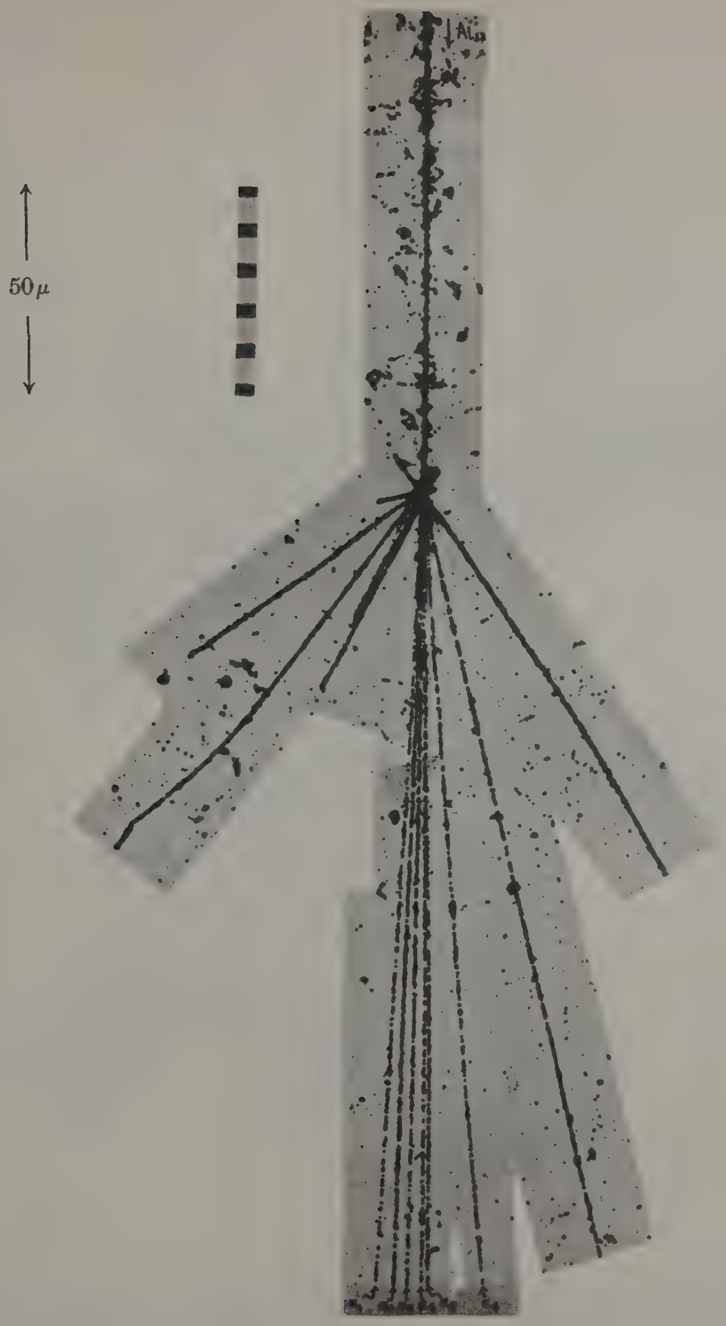


PLATE I.

A nucleus of charge Z=13±1 with a velocity $\beta=0{\cdot}55$ suffers a collision in the emulsion and gives rise to a "jet" of six α-particles (considered to be components of the primary particle) and low energy disintegration fragments of the target nucleus. Observer: Miss J. HERAPER.

To face page 966

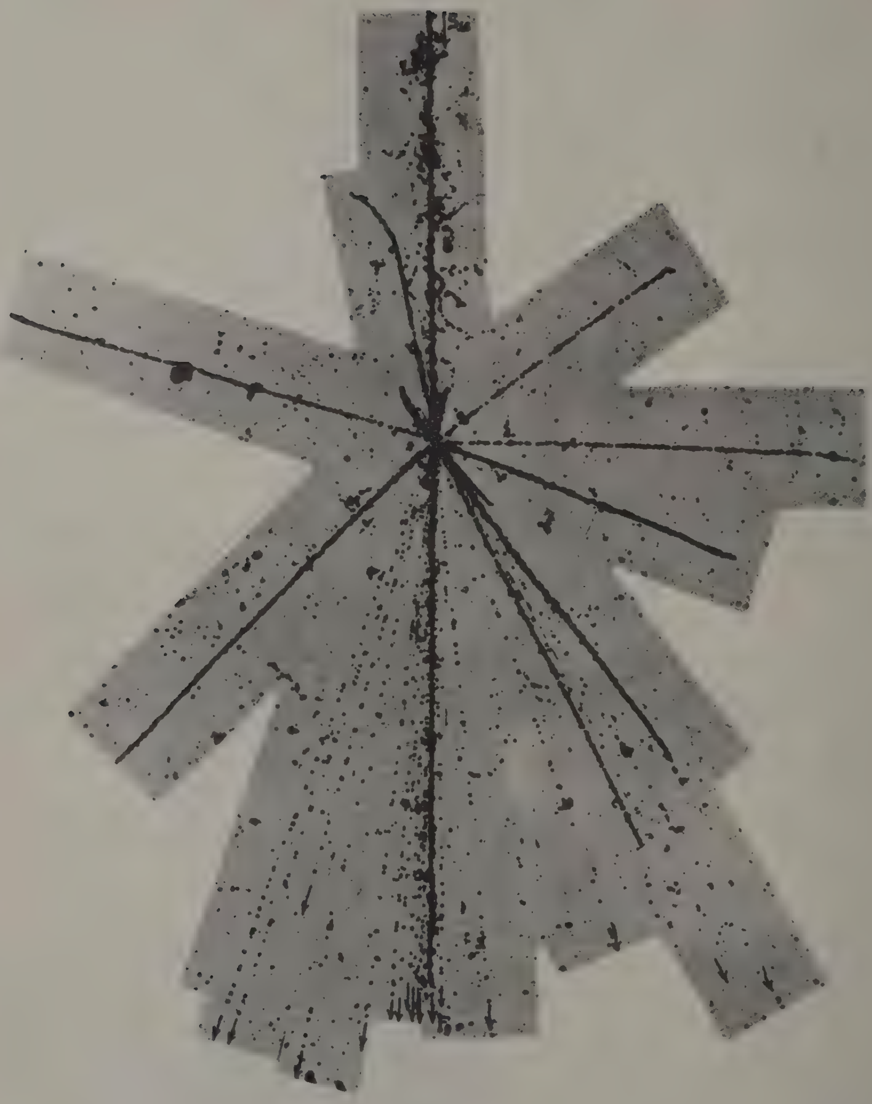


PLATE II.

A fast nucleus of charge Z=16±1 interacts with one of silver or bromine in the emulsion. The nucleus emerging from the encounter has a charge Z=9±1 and is accompanied by twenty-five singly charged particles moving at relativistic velocities, and in directions contained by a cone of semi-vertical angle 30°. The tracks of fifteen of these particles have been reproduced in this photomicrograph. Observer: Miss E. JAMES.

The calculations have been made using the theoretical equation for protons given by Livingston and Bethe (1937)

$$-\frac{dE}{dx}=\frac{4\pi Ne^4}{mv^2}\left[Z'(\ln\frac{2mv^2}{I(1-\beta^2)}-\beta^2)-C_K\right],$$

where $-dE/dx$ is the rate of loss of energy, Z' the effective number of electrons per atom, I the mean excitation energy of the atom of the absorber and C_K a correction term.

For elements of low atomic number, Z' is taken as the total number of electrons, Z, and the non-excitation of the inner shell electrons for low proton energies is corrected by the term C_K. Values of C_K are given by Livingston and Bethe (1937) in terms of the energy of the proton and a critical energy depending on the ionization potential of the electrons. For elements of high atomic number Z' is taken as the effective number of electrons (Hönl 1933).

The value of I is usually chosen to give agreement with experimental values for the energy loss at low energies. On the basis of the Fermi Thomas model of the atom, Bloch (1933) showed that approximately a linear relationship should hold between I and Z for elements of high atomic number. The constant of proportionality, α, given by Wheller (1941), is 11·5 eV. while Mano (1934 a, b) has given the value of 10·5 eV. for elements with Z greater than 20 but shows that α fluctuates for Z less than 20. The values adopted by Livingston and Bethe (1937) for hydrogen, helium, aluminium, air and gold differ slightly from those of Mano, but the variation of α with Z follows the same pattern.

Before the computation of the range energy relation in carbon can be carried out, an appropriate value of I must be found. In view of the variation of α for elements of low atomic weight, it was decided to find a value which gave an atomic stopping power relative to aluminium in agreement with experimental values, rather than take the theoretical value of 11·5 eV. Using for Z' the total number of electrons both for aluminium and carbon and a value of I for aluminium as 150 eV., the value of I for carbon was found to be 50·3 eV., and hence α is 8·4 eV. This value seems to fit in with the pattern of variation of α with Z. Similarly for silver and copper the calculated values of I were 365 eV. for silver and 293 eV. for copper, using values of Z' 39·6 and 27·4 respectively.

Using the value of I as 50·3 eV. for carbon and the appropriate values of C_K, dE/dx was calculated at 20 MeV. intervals up to 200 MeV. Below 20 MeV. the known experimental values were taken. The range was then computed by numerical integration and is given in Table I.

These values for the range have been used as a measure of the energy of protons scattered into a triple coincidence telescope from a polythene radiator bombarded by neutrons ejected from beryllium and carbon targets. The end point of the spectrum of the scattered protons was compared with that using aluminium for the absorber. The energy in this case was deduced from the range energy curves of Smith (1947).

TABLE I.

Energy MeV.	$\frac{1}{\rho}\frac{dE}{dx}$ MeV./gm. cm.$^{-2}$	Range gm. cm.$^{-2}$
2·5	132·5	0·0113
5	77·3	0·0374
10	44·15	0·127
15	31·58	0·264
20	24·96	0·442
40	14·16	1·55
60	10·23	3·24
80	8·16	5·45
100	6·89	8·13
120	6·03	11·24
140	5·39	14·76
160	4·91	18·65
180	4·54	22·89
200	4·23	27·45

The energies so determined agree with each other to 2 MeV. and to within 2 MeV. of the value expected from the theoretical energy of the neutrons, the experimental error being also about 2 MeV.

August 2nd, 1950. A. E. TAYLOR.

REFERENCES.

BLOCH, F., 1933, *Zeits f. Phys.*, **81,** 363.
HÖNL, H., 1933, *Zeits. f. Phys.*, **84,** 1.
LIVINGSTON, M. S., and BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9,** 261.
MANO, G., 1934 a, *Ann. Phys., Lpz.*, **1,** 407.; 1934 b, *J. Phys. Radium*, **5,** 628.
SMITH, J. H., 1947, *Phys. Rev.*, **71,** 32.
TAYLOR, A. E., PICKAVANCE, T. G., CASSELS, J. M., and RANDLE, T. C., 1950, *Nature, Lond.*, **165,** 4207, 967–968.
WHELLER, J. A., 1941, *Phys. Rev.*, **60,** 754.

[*The Editors do not hold themselves responsible for the views expresesd by their correspondents.*]

LXXXIX. *The Statistics of Correlated Events.*—I.

By C. DOMB,
Clarendon Laboratory, Oxford*.

[Received, June 20, 1950].

SUMMARY.

The present paper is an attempt to formulate the mathematical properties of events occurring in time with a correlation between them. One effect of the correlation is a change in the mean square deviation of the number of events occurring in an interval of time t. By analogy with the shot effect in the presence of space change, this is denoted by a factor Γ^2 for large t, and Γ^2 is evaluated in terms of the interval distributions between events. For finite intervals t a corresponding factor $\gamma^2(t)$ is introduced, which usually tends to 1 as $t \to 0$. A correlation function between events occurring at times separated by an interval τ is defined, and its relation to $\gamma^2(t)$ is discussed. A generalization of Campbell's Theorem applying to correlated events is derived. Finally the problem of the random partitioning of correlated events is discussed.

§1. INTRODUCTION.

SUPPOSE that events occur randomly and independently in time, the probability of an event occurring in the interval $(t, t+dt)$ being $\lambda\, dt$; such events will be referred to briefly as " random events ". There are a number of problems in which the concept of random events can be regarded as a valid physical approximation, and as a typical example we may take the emission of electrons from a cathode in a valve. The properties of random events have been studied extensively, particularly in connection with the shot effect and the properties of random noise. Little attention, however, seems to have been paid to the properties of " correlated events ", that is, events in which the probability of an occurrence at a certain time is dependent on the distribution of events which have occurred previously.

The most common physical problem in which there is a correlation between events is the shot effect in the presence of space charge, and here the correlating mechanism is enormously complicated. However, the essential physical results of the correlation can be expressed quite simply. For random events it can readily be shown (*e. g.* Bateman 1910) that, if n is the number of events occurring in a finite interval of length t,

$$E(n)=\lambda t, \quad \ldots \ldots \ldots \quad (1)$$

$$E(n^2)-[E(n)]^2=\lambda t, \quad \ldots \ldots \ldots \quad (2)$$

---

* Communicated by Prof. M. H. L. Pryce.

where $E(u)$ denotes the expectation, or mean value of u. In the presence of space charge there are two changes in the above formulæ. Firstly, the mean number of events is reduced, so that we now have

$$E(n)=\alpha\lambda t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where α lies between 0 and 1. Secondly, the fluctuations are smoothed, so that

$$E(n^2)-[E(n)]^2 \sim \Gamma^2\alpha\lambda t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where Γ^2 lies between 0 and 1, and is usually referred to as the space charge reduction factor. It should be noted that although (2) is valid for an interval of any length, t, the definition of Γ^2 in (4) usually refers to the asymptotic value for large t; this point will be considered in more detail later in § 5.

A good deal of work has been done on the evaluation of Γ^2 in particular cases of physical interest (see North (1940) for references). In point of fact, however, the relations (3) and (4) arise immediately a correlation is introduced between events, and a fundamental analysis in terms of the theory of probability should be possible, giving α and Γ^2 in terms of the correlations between events. As far as the writer is aware, the only attempt along these lines is that of Whittaker (1938).

In the present paper it is proposed to consider classes of correlated events in which the probability of an event occurring at a given time depends on the time of occurrence of the previous r events. This means that any r intervals between successive occurrences are correlated, and such events will be referred to as events of class r. For many types of non-random events this is not a particularly simple definition. For example, if we consider a purely random primary series in which each primary event is followed by a secondary event at a fixed time τ after it, one can easily see that the complete set of primary and secondary events are of class ∞, even though the properties of the set are easy to discuss by elementary means. Nevertheless, given any set of events, the property immediately accessible to observation is the length of successive intervals, and one of our purposes is to relate the interval distributions to α and Γ^2 defined in (3) and (4).

§ 2. EVENTS OF CLASS 1.

We first consider in detail the properties of events in which the probability of an occurrence depends only on the time which has elapsed from the previous occurrence. This type of problem arises in the theory of particle counters, and has been discussed by a number of authors (see *e. g.* Feller 1948); it has also arisen recently in connection with genetics (Owen 1949). The advantageous feature of events of class 1 is that an exact solution is available in terms of Laplace transforms (Domb 1948, Owen 1949), and the physical properties of the solution can readily be determined. We now proceed to deduce this solution; much of the development will parallel previous discussions, but the results are required in a somewhat different form.

Suppose that if no event has occurred in an interval ξ after a given event, the probability of an occurrence in the interval $(\xi,\ \xi+d\xi)$ is $\lambda f(\xi)d\xi$. In cases of practical interest, we should expect $f(\xi)$ to tend to unity for large ξ. Let $\phi(\xi)$ be the probability that no event occurs in an interval ξ after a given event. Then

$$\phi(\xi+d\xi)=\phi(\xi)[1-\lambda f(\xi)d\xi]$$

or

$$\phi(\xi)=e^{-\lambda\mathrm{F}(\xi)}\left(\mathrm{F}(\xi)=\int_0^\xi f(u)\,du\right). \quad \text{. . . . . (5)}$$

The probability that the interval between two successive events has length between ξ and $\xi+d\xi$ is thus

$$\lambda f(\xi)e^{-\lambda\mathrm{F}(\xi)}d\xi=g(\xi)d\xi. \quad \text{. . . . . (6)}$$

Now let $u_n(t)$ be the probability that n events occur in the interval $(0,t)$. We can divide $u_n(t)$ into mutually exclusive classes in which the last event occurs in $(t-\eta;\ t-\eta+d\eta)$, the probability being $u_n(t,\eta)d\eta$. It is easy to see that

$$u_n(t,\eta)=e^{-\lambda\mathrm{F}(\eta)}u_n(t-\eta,0)=e^{-\lambda\mathrm{F}(\eta)}w_n(t-\eta), \quad \text{. . . (7)}$$

so that

$$u_n(t)=\int_0^t u_n(t,\eta)\,d\eta=\int_0^t e^{-\lambda\mathrm{F}(\eta)}w_n(t-\eta)\,d\eta. \quad \text{. . . (8)}$$

We may rewrite (8) symbolically in the form

$$u_n(t)=\mathrm{P}w_n(t), \quad \text{. . . . . (9)}$$

where P is the linear operator defined by (8).

$w_n(t)$ can similarly be divided into mutually exclusive classes in which the last event but one occurs in $(t-\zeta,\ t-\zeta+d\zeta)$, and as above, we deduce that

$$w_n(t)=\int_0^t w_{n-1}(t-\zeta)g(\zeta)\,d\zeta, \quad \text{. . . . . . . (10)}$$

or symbolically

$$w_n(t)=\mathrm{Q}w_{n-1}(t)=\mathrm{Q}^{n-1}w_1(t), \quad \text{. . . . . . . (11)}$$

where Q is the linear operator defined by (10).

In discussing the probability distribution of numbers of events it is usually convenient to introduce the moment generating function, for example,

$$w(\theta,\ t)=\sum_1^\infty w_n(t)e^{n\theta}=\sum_1^\infty e^{n\theta}\mathrm{Q}^{n-1}w_1(t). \quad \text{. . . . . (12)}$$

Now for any finite t we can reasonably assume that the probability of n events occurring tends to zero as n tends to infinity. Thus we may write symbolically

$$w(\theta,t)=\frac{w_1(t)e^\theta}{1-\mathrm{Q}e^\theta}, \quad \text{. . . . . (13)}$$

so that $w(\theta, t)$ satisfies an integral equation. We have discussed the solution in terms of linear operators, since the whole development can be formally generalized to events of class r, provided the operators P and Q are re-defined appropriately.

We now particularize to events of class 1, and take Laplace transforms* in t. The Laplace transform of any function will be denoted by a capital letter; for example $G(p)$ will be used for the Laplace transform of $g(t)$ defined in (6). The operators P, Q now have a very simple interpretation, Q corresponding to a multiplication by $G(p)$, and P to a multiplication by $[1-G(p)]/p$. Equation (13) becomes

$$w(\theta, p)=\frac{w_1(p)e^{\theta}}{1-G(p)e^{\theta}}=\frac{w_1(p)}{e^{-\theta}-G(p)} \quad . \; . \; . \; . \; . \quad (14)$$

and correspondingly,

$$U(\theta, p)=\frac{w_1(p)[1-G(p)]}{p[e^{-\theta}-G(p)]}. \quad . \; . \; . \; . \; . \; . \; . \quad (15)$$

To specify the solution completely we should now consider the initial conditions at $t=0$, and hence determine $w_1(t)$. But in fact, as we should expect, the asymptotic solution for sufficiently large t is independent of $w_1(t)$ and depends only on the form of $g(t)$. This asymptotic solution can be derived by expanding (15) as a series of residues (Domb 1948),

$$u(\theta, t)=\sum_s A_s(\theta, t)e^{m_s t}, \quad . \; . \; . \; . \; . \; . \; . \quad (16)$$

where the m_s are the roots of

$$e^{-\theta}-G(p)=0. \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (17)$$

The dominant term is given by the real root m_1 of (17), and one can readily obtain an expansion of this root in powers of θ, the coefficients of which are the asymptotic values of the cumulants of the $u_n(t)$ distribution. We find that

$$\left.\begin{aligned} E(n)&\sim t/g_1, \\ E(n^2)-[E(n)]^2&\sim (g_2/g_1^3-1/g_1)t, \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \quad (18)$$

where
$$g_r=\int_0^\infty t^r g(t)\,dt,$$

and is the rth moment of the $g(t)$ distribution. Hence in the notation of §1,

$$\left.\begin{aligned} \alpha&=1/\lambda g_1, \\ \Gamma^2&=g_2/g_1^2-1. \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (19)$$

---

* We depart here from the notation of Domb (1948) and define the Laplace transform of a function $y(t)$ as $\int_0^\infty y(t)e^{-pt}\,dt$. This is now adopted by many writers, and leads to some simplication of formulæ.

It is important to observe that Γ^2 depends only on the first two moments of the $g(t)$ distribution. Thus $g(t)$ could be replaced by any other distribution having the same first two moments without altering Γ^2; as a particular example $g(t)$ could be formally replaced by an error distribution whose first two moments are g_1 and g_2, although this is not feasible physically, since t cannot be negative. Similar considerations apply to higher moments of the $u_n(t)$ and $g(t)$ distributions, but these are not usually of physical interest.

· We shall now give a few practical examples to illustrate the evaluation of (19).

(i) An α-particle counter of constant recovery time β. $g(t)$ is the sum of two distributions, a fixed length β and a distribution $\lambda e^{-\lambda u}du$.

$$\left.\begin{array}{ll} g_1=\beta+1/\lambda, & \alpha=1/(1+\lambda\beta), \\ g_2-g_1^2=1/\lambda^2, & \Gamma^2=1/(1+\lambda\beta)^2, \end{array}\right\}, \quad . \quad . \quad . \quad (20)$$

so that Γ^2 is always less than 1, as we should expect.

(ii) Consider $g(t)=\dfrac{\lambda^\nu t^{\nu-1}}{\Gamma(\nu)}e^{-\lambda t}$; $\nu=1$ corresponds to the uncorrelated case, $0<\nu<1$ corresponds to an attractive correlation, and $\nu>1$ to a repulsive correlation. In this case $G(p)=(\lambda/p+\lambda)^\nu$, and equation (17) can be solved exactly giving

$$m_1=-\lambda+\lambda e^{\theta/\nu}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

(There are no other roots of (17) in this case, the remainder being given by an integral round the branch point $p=-\lambda$). We readily deduce that

$$\alpha=1/\nu, \qquad \Gamma^2=1/\nu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

so that $\Gamma^2<1$ for repulsive correlation, and >1 for attractive correlation.

(iii) Let $g(t)=\beta\delta(t)+(1-\beta)\lambda e^{-\lambda t}$; this means that there are finite probabilities β, β^2 . . . . of the simultaneous emission of 2, 3, events. Then

$$\alpha=1/1-\beta, \qquad \Gamma^2=1+\beta/1-\beta. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

This is a particular case of a more general problem considered by Whittaker (1938).

§ 3. Formal Generalization to Events of Class *r*.

We now suppose that if the intervals between the last r occurrences are $\xi_1, \xi_2, \ldots \xi_{r-1}$ the probability that an event occurs in the interval $[\eta, \eta+d\eta]$ from the last occurrence is $\lambda f(\xi_1, \xi_2 \ldots \xi_{r-1}; \eta)$ (this is a conditional probability). The relations (5) to (13) can be formally generalized, many of them now assuming the form of conditional probabilities. Thus

$$\left.\begin{array}{l} \phi(\xi_1, \ldots \xi_{r-1}; \eta)=e^{-\lambda F(\xi_1, \ldots \xi_{r-1}; \eta)}, \\ \left[F(\xi_1, \ldots \xi_{r-1}; \eta)]=\int_0^\eta f(\xi_1, \ldots \xi_{r-1}; u)\,du\right]. \end{array}\right\} \quad . \quad (24)$$

and the probability $h(\xi_1, \dots \xi_{r-1}; \eta)$ that the interval following a given event lies between η and $\eta+d\eta$, when the previous $(r-1)$ intervals are $\xi_1, \dots \xi_{r-1}$ is

$$\lambda f(\xi_1, \dots \xi_{r-1}; \eta) e^{-\lambda F(\xi_1, \dots \xi_{r-1}; \eta)} d\eta. \quad . \quad . \quad . \quad . \quad . \qquad (25)$$

As before $u_n(t)$ can be divided into mutually exclusive classes

$$u_n(t; \xi_1, \dots \xi_{r-1}; \eta)$$

and we can define $w_n(t; \xi_1, \dots \xi_{r-1})$ $[=u_n(t; \xi_1, \dots \xi_{r-1}; 0]$. The operators P and Q are now given by

$$\left.\begin{aligned} \mathrm{P}w_n(t; \xi_1, \dots \xi_{r-1}) &= \int_0^\infty d\xi_1 \dots \int_0^\infty d\xi_{r-1} \int_0^t d\eta e^{-\lambda F(\xi_1, \dots \xi_{r-1}; \eta)} \\ &\qquad \times w_n(t-\eta, \xi_1, \dots \xi_{r-1}), \\ \mathrm{Q}w_n(t; \xi_1, \dots \xi_{r-1}) &= \int_0^\infty h(\eta, \xi_1, \dots; \xi_{r-1})\, d\eta \\ &\qquad \times w_n(t-\xi_{r-1}, \eta, \xi_1 \dots \xi_{r-2})\, d\eta. \end{aligned}\right\} \quad . \quad . \quad . \qquad (26)$$

Equation (13) now furnishes an integral equation for $w(\theta, \xi_1, \dots \xi_{r-1})$, the operator Q being given by (26).

Unfortunately no transformation seems to be available in the general case to provide the analogue of the Laplace transform solution for events of class 1. On physical grounds we might expect the solutions to have a fairly simple asymptotic character, and we shall find this borne out in the next section.

§ 4. REFORMULATION IN TERMS OF INTERVAL DISTRIBUTIONS.

For event of class r any r adjacent intervals are correlated, but any intervals separated by r or more intervening intervals are uncorrelated. Let $t_n(\xi)d\xi$ be the probability that the sum of n adjacent intervals lies between ξ and $\xi+d\xi$. Then the probability that n events or more occur in a time t after a given occurrence is given by

$$q_n(t) = \int_0^t t_n(\xi)\, d\xi \qquad (n \geqslant 1). \quad . \quad . \quad . \quad . \quad . \qquad (27)$$

Thus the probability that exactly n events occur is

$$p_n(t) = q_n(t) - q_{n+1}(t) \qquad (n \geqslant 1). \quad . \quad . \quad . \quad . \quad . \quad . \qquad (28)$$

Hence the problem of determining the distribution of n is equivalent to determining $t_n(\xi)$.

The explicit evaluation of $t_n(\xi)$ is complicated; for large n we should expect $t_n(\xi)$ to be normally distributed, and it might be possible to derive asymptotic results on this basis. However, we shall find it more convenient to derive the moments of the $t_n(\xi)$ distribution (which can be done quite readily even for events of class r) and deduce results by analogy with events of class 1, for which an exact solution is available. For events of class 1, the cumulants of the $t_n(\xi)$ distribution are simply n times the cumulants of the $g(\xi)$ distribution, so that

$$\left.\begin{aligned} &\mu_1^{(n)}=ng_1,\\ &\mu_2^{(n)}-\mu_1^{(n)^2}=n(g_2-g_1^2). \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \; . \qquad (29)$$

We have seen that α and Γ^2 depend only on the two moments g_1 and g_2, and hence only on the two terms of (29). For events of class r let

$$g(\xi_1, \xi_2 \ldots \xi_r)\, d\xi_1 \ldots d\xi_r$$

be the joint probability distribution of r adjacent intervals, and denote the mean value of any interval, $\mathrm{E}(\xi_1)$, by g_1, and $\mathrm{E}(\xi_1\ \xi_i)$ by g_{1i}; thus if $i>r$, $\mathrm{E}(\xi_1\ \xi_i)=g_1^2$. Then

$$\mu_1^{(n)}=\mathrm{E}(\xi_1+\ldots+\xi_n)=\mathrm{E}(\xi_1)+\mathrm{E}(\xi_2)+\ldots\mathrm{E}(\xi_n)=ng_1 \quad . \qquad (30)$$

$$\begin{aligned} \mu_2^{(n)}&=\mathrm{E}\left(\sum_{i=1}^{n}\xi_i\right)^2=\mathrm{E}\left(\sum_{i=1}^{n}\xi_i^2+2\sum_{i<j}\xi_i\xi_j\right)\\ &=ng_{11}+2(n-1)g_{12}+2(n-2)g_{13}+\ldots 2(n-r+1)g_{1r}\\ &\qquad +\{n^2-n-2[(n-1)+\ldots(n-r+1)]\}g_1^2. \quad . \; . \; . \qquad (31) \end{aligned}$$

Hence

$$\begin{aligned} \mu_2^{(n)}-\mu_1^{(n)^2}&=n[g_{11}+2g_{12}+\ldots 2g_{1r}-\overline{2r-1}g_1^2]\\ &\qquad -2[g_{12}+2g_{13}+\ldots(r-1)g_{1r}-\frac{r(r-1)}{2}g_1^2]. \quad . \; . \qquad (32) \end{aligned}$$

The second term of (32) does not involve n, and since r is a finite number, it will become negligible for sufficiently large n; in fact, it represents something in the nature of an end effect, and we should not expect it to influence the asymptotic distribution of the n's. With the exception of this term there is a complete analogy between (20) and (32), and (29), and the problem of class r is thus equivalent to a problem of class 1, in which g_2 is replaced by

$$g_{11}+2(g_{12}+\ldots g_{1r})-(2r-2)g_1^2. \quad . \; . \; . \; . \qquad (33)$$

Using the result (19) for events of class 1, we deduce that, for events of class r,

$$\left.\begin{aligned} &\alpha=1/\lambda g_1,\\ &\Gamma^2=\frac{g_{11}+2g_{12}+\ldots 2g_{1r}-(2r-1)g_1^2}{g_1^2}. \end{aligned}\right\} \quad . \; . \; . \; . \qquad (34)$$

We may also observe that there would be no difficulty in deriving higher moments in a similar manner.

§ 5. THE CORRELATION FUNCTION.

So far we have been dealing with asymptotic values for large intervals of time; we shall now consider an interval $[0, t]$ of finite length. Let us divide the interval $[0, t]$ into a large number N of sub-intervals of length δt, and denote the number of events occurring in the ith sub-interval, $[t_i\, t_i+\delta t]$, by δn_i. Then, since we have assumed that the process is in a "stationary" condition, E (δn_i) is independent of t_i, and E $(\delta n_i\, \delta n_j)$ depends only on (t_j-t_i) Therefore

$$E(n)=E\left[\sum_{i=1}^{N}\delta n_i\right]=\sum_{i=1}^{N}E(\delta n_i)=\lambda\alpha t, \quad . \quad . \quad . \quad . \quad (35)$$

so that this relation is the same for finite t. We now write

$$E(\delta n_i\delta n_j)=(\lambda\alpha\delta t)^2[1+\rho(t_i-t_j)], \quad . \quad . \quad . \quad . \quad . \quad (36)$$

where $(\rho\tau)$ can be regarded as a correlation function, indicating the time during which correlation is appreciable. Then

$$E(n^2)=E\left(\sum_{i=1}^{N}\delta n_i\right)^2=E\left(\sum_{ij=1}^{N}\delta n_i\delta n_j\right)=\sum_{ij=1}^{N}E(\delta n_i\delta n_j), \quad . \quad (37)$$

and as N tends to ∞, this can readily be transformed into

$$2(\lambda\alpha)^2\int_0^t(t-\tau)[1+\rho(\tau)]d\tau=\lambda^2\alpha^2t^2+2(\lambda\alpha)^2\int_0^t(t-\tau)\rho(\tau)d\tau. \quad . \quad (38)$$

Hence if we define $\gamma^2(t)$ by the relation

$$E(n^2)-[E(n)]^2=\gamma^2\alpha\lambda t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

for a finite interval t, we shall have

$$\gamma^2(t)=\frac{2\lambda\alpha}{t}\int_0^t(t-\tau)\rho(\tau)d\tau. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

(It should be noted that $\rho(\tau)$ has a δ-function singularity at $\tau=0$ corresponding to $i=j$ in (36)).

It is not difficult to see that, if the probability of two events occurring in an interval δt is $0(\delta t^2)$, $\gamma^2(t)$ must tend to unity as $t \to 0$. On the other hand by definition $\gamma^2(t)\to\Gamma^2$ as $t\to\infty$, so that $\gamma^2(t)$ passes from 1 to Γ^2 as t goes from 0 to ∞. Letting $t\to\infty$ in (40),

$$\Gamma^2=2\lambda\alpha\int_0^\infty\rho(\tau)d\tau. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

The above formulæ are well illustrated by the example considered previously of primary and secondary events separated by a fixed interval β, for which

$$2\lambda\alpha\rho(\tau)=\delta(\tau)+\delta(\tau-\beta). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Here,

$$\gamma^2(t)=\begin{cases}1 & t<\beta,\\ 2 & t>\beta,\end{cases}$$

so that for $t<\beta$, the events may be regarded as single and random, but for $t>\beta$ they are to be regarded as doublets.

We can readily make use of the correlation function to determine the correlation between any two finite intervals of lengths t_1 and t_2, separated by an interval of length u. If m refers to the number of events occurring in the first interval, and n to the number of events occurring in the second,

$$E(mn)=E\left[\left(\sum_{i=1}^{M}\delta m_i\right)\left(\sum_{j=1}^{N}\delta n_j\right)\right]=\sum_{i=1}^{M}\sum_{j=1}^{N}E(\delta m_i\delta n_j\Big),$$

and this can readily be transformed into

$$\lambda^2\alpha^2\Bigg[t_1t_2+\int_0^{t_1}\tau\rho(u+\tau)d\tau+t_1\int_{t_1}^{t_2}\rho(u+\tau)\,d\tau \\ +\int_{t_2}^{t_1+t_2}(t_2+t_1-\tau)\rho(u+\tau)d\tau\Bigg]\,(t_2\geqslant t_1). \quad . \quad . \quad . \quad (43)$$

For events of class 1, we can obtain an explicit solution for the correlation function by means of Laplace transforms. We require first to specify the value of $W_1(p)$ in (15) to correspond to the " stationary " condition. This can most readily be determined by the condition that $E(n)$ must be equal to $\lambda\alpha t$ for any finite t: hence $W_1(p)=[1-G(p)]/g_1p$. (An alternative derivation is given by Domb (1948)). We therefore deduce from (5) that the Laplace transform of $E(n^2)$ is

$$\frac{1+G(p)}{g_1p^2[1-G(p)]}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Let $R(p)$ be the Laplace transform of $\rho(t)$. Then taking Laplace transforms in (38) and equating to (44) we have

$$\frac{2(\lambda\alpha)^2}{p^3}+\frac{2(\lambda\alpha)^2R(p)}{p^2}=\frac{\lambda\alpha[1+G(p)]}{p^2[1-G(p)]}$$

or

$$R(p)=\frac{1+G(p)}{2\lambda\alpha[1-G(p)]}-\frac{1}{p}=\frac{1}{2\lambda\alpha}-\frac{1}{p}+\frac{G(p)}{\lambda\alpha[1-G(p)]}. \quad . \quad . \quad (45)$$

The first term of the final expression in (45) corresponds to the δ-function singularity at $t=0$. When there is no correlation, so that $G(p)=\lambda/p+\lambda$ the remaining terms in (45) vanish, as expected.

We now apply (45) to some particular cases considered previously in §2.

(i) For an α-particle counter of constant recovery time β

$$G(p)=\lambda e^{-\beta p}/(p+\lambda) \qquad \text{and} \qquad \alpha=1/(1+\beta\lambda).$$

Therefore

$$R(p)-\frac{1}{2\lambda\alpha}=\frac{(1+\beta\lambda)e^{-\beta p}}{p+\lambda-\lambda e^{-\beta p}}-\frac{1}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

and this can be expanded in the form

$$-\frac{1}{p}+\frac{(1+\beta\lambda)e^{-\beta p}}{p+\lambda}+\frac{\lambda(1+\beta\lambda)e^{-2\beta p}}{(p+\lambda)^2}+\ldots. \quad . \quad . \quad . \quad (47)$$

Thus we have an explicit expression for $\rho(t)$ in the form

$$\rho(t)-\frac{1}{2\lambda\alpha}\delta(t)$$
$$=-1+(1+\beta\lambda)\left[e^{-\lambda t_1}+\lambda t_2 e^{-\lambda t_2}+\frac{\lambda^2 t_3^2}{2\,!}e^{-\lambda t_3}+\ \ldots\right](t_r=t-r\tau), \quad . \quad (48)$$

where, if $(r-1)\tau\leqslant t\leqslant r\tau$, r terms are to be taken into account.

It is also of interest to consider the asymptotic behaviour of $\rho(t)$, and this can readily be obtained from the poles of (46). There is no pole at $p=0$, and hence the dominating term in an asymptotic series such as (16) is given by the real root, $-\eta$, of

$$p+\lambda-\lambda e^{-\beta p}=0.$$

Then
$$\rho(t)\sim\frac{(1+\beta\lambda)(1-\eta/\lambda)}{1+\beta(\lambda-\eta)}e^{-\eta t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

(ii) Let

$$g(t)=\frac{\lambda^\nu t^{\nu-1}}{\Gamma(\nu)}e^{-\lambda t}, \qquad \text{so that} \qquad G(p)=\left(\frac{\lambda}{p+\lambda}\right)^\nu \qquad \text{and} \qquad g_1=\nu/\lambda.$$

We have
$$R(p)-\frac{\nu}{2\lambda}=\frac{\nu\lambda^{\nu-1}}{(p+\lambda)^\nu-\lambda^\nu}-\frac{1}{p}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

When ν is an integer, r, this can be readily evaluated explicitly as a sum of exponentials, the poles being given by $p/\lambda=-1+e^{2s\pi i/r}$; for example when $\nu=2$,

$$\rho(t)-\frac{\delta(t)}{\lambda}=-e^{-2\lambda t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

The asymptotic form of $\rho(t)$ for general ν cannot be obtained as readily as in the previous example, since branch line integrals must be considered.

§ 6. THE GENERALIZATION OF CAMPBELL'S THEOREM.

For purely random events Campbell considered the distribution of amplitude in an instrument, when the effect of each event could be represented by an amplitude $s(t)$ after time t. We shall now consider the generalization of his results to correlated events. Let y be the amplitude at the time of observation, and let us measure time t backwards from this point; again we divide our time interval into a large number of small sub-intervals $[t_1, t_1+\delta t]$, and

$$y=\underset{\delta t\to 0}{\mathrm{Lt}}\ \underset{i}{\Sigma}\,\delta n_i\, s(t_i). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

Therefore
$$E(y)=\underset{\delta t\to 0}{\mathrm{Lt}}\ \underset{i}{\Sigma}E(\delta n_i)s(t_i)=\lambda\alpha\int_0^\infty s(t)\,dt \quad . \quad . \quad . \quad . \quad (53)$$

and
$$E(y^2)=\underset{\delta t\to 0}{\mathrm{Lt}}\ \underset{i}{\Sigma}\Sigma\,E(\delta n_i\delta n_j)s(t_i)s(t_j). \quad . \quad . \quad . \quad . \quad . \quad (54)$$

Substituting from (36) and proceeding as before, this double integral can be transformed into

$$(\lambda\alpha)^2\left[\int_0^\infty s(t)\,dt\right]^2+2(\lambda\alpha)^2\int_0^\infty d\tau\rho(\tau)\int_0^\infty s(t)s(t+\tau)\,dt \quad . \quad . \quad (55)$$

and thus

$$\mathrm{E}(y^2)-[\mathrm{E}(y)]^2=2(\lambda\alpha)^2\int_0^\infty \rho(\tau)d\tau\int_0^\infty s(t)s(t+\tau)\,dt. \quad . \quad . \quad (56)$$

This may be regarded as a generalized form of Campbell's theorem, the δ-function term in $\rho(\tau)$ giving rise to the usual term $\int_0^\infty s^2(t)\,dt$. As a particular example, if $\rho(t)$ is only significant during a time for which $s(t)$ does not change appreciably, the second integral in (56) can be replaced by $\int_0^\infty s^2(t)dt$, and using (41) we have

$$\mathrm{E}(y^2)-[\mathrm{E}(y)]^2=\lambda\alpha\Gamma^2\int_0^\infty s^2(t)\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

This is the usual form of the theorem for the shot effect in the presence of space charge.

If we now take $s(t)$ to correspond to a sharply tuned circuit at frequency $\omega/2\pi$ $\mathrm{E}(y^2)$ will correspond to a "power spectrum", and we obtain the standard formula

$$\mathrm{P}(\omega)=\int_0^\infty 2\rho(\tau)\cos\omega\tau\,d\tau. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

This is especially useful in considering events of class 1, since $\mathrm{R}(p)$ has been evaluated explicitly in (45), and

$$\mathrm{P}(\omega)=\mathrm{R}(i\omega)+\mathrm{R}(-i\omega). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

As a particular case, for the α-particle counter we deduce from (46) that

$$\mathrm{P}(\omega)=\frac{1+\beta\lambda}{\lambda}\left[1-\frac{1-\cos\beta\omega+\frac{\omega}{\lambda}\sin\beta\omega}{1-\cos\beta\omega+\frac{\omega}{\lambda}\sin\beta\omega+\omega^2/2\lambda^2}\right]. \quad . \quad . \quad (60)$$

It is interesting to note that as λ tends to ∞, this tends to zero for all frequencies other than multiples of $2\pi/\beta$, which is in accordance with what we would expect physically.

A rather interesting example is that of primary and secondary events separated by a fixed interval β. From (42) we readily deduce that

$$\mathrm{P}(\omega)=\frac{1}{\lambda}\cos^2\frac{\beta\omega}{2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

§7. THE PARTITIONING OF EVENTS

We now consider the effect of the random partitioning of correlated events into two groups. This problem has a physical application to valves containing two or more electrodes, and gives rise to the phenomenon of "partition noise." Let α_1 and α_2 be the probabilities that any given event falls into groups 1 or 2 respectively $(\alpha_1+\alpha_2=1)$. Then the probability of an occurrence of n_1 events in group 1 and n_2 in group 2 in an interval $[0, t]$ is

$$\psi(n_1, n_2\,;\ t)=\frac{n\,!}{n_1\,!\ n_2\,!}\alpha_1^{n_1}\alpha_2^{n_2}u_n(t), \quad . \quad . \quad . \quad . \quad . \quad (62)$$

where $n=n_1+n_2$ and $u_n(t)$ is the probability of n events occurring in $[0, t]$. We have seen from §'s 2, 3 that $u_n(t)$ may be formally written as $PQ^{n-1}w_1(t)$ where P and Q are certain linear operators. The joint moment generating function of n_1 and n_2 is given by

$$\begin{aligned}\Psi(\theta_1, \theta_2\,;\ t)&=\sum_{n_1n_2}\psi(n_1, n_2\,;\ t)e^{(n_1\theta_1+n_2\theta_2)}\\ &=PQ^{-1}\sum_{n_1n_2}\frac{(n_1+n_2)\,!}{n_1\,!\ n_2\,!}(\alpha_1Qe^{\theta_1})^{n_1}(\alpha_2Qe^{\theta_2})^{n_2}w_1(t)\\ &=P\sum_{n}(\alpha_1e^{\theta_1}+\alpha_2e^{\theta_2})^nQ^{n-1}w_1(t). \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)\end{aligned}$$

This is identical in form with the moment generating function for n except that e^{θ} is replaced by $\alpha_1e^{\theta_1}+\alpha_2e^{\theta_2}$. Hence if we know the expansion of the generating function for n we can immediately determine the corresponding expansion for n_1 and n_2 by replacing θ by

$$\ln\left[1+\alpha_1(e^{\theta_1}-1)+\alpha_2(e^{\theta_2}-1)\right].$$

In particular if we take into account only two terms of the cumulant generating function of n, we have

$$\ln u(\theta, t)\sim\lambda\alpha t(\theta+\Gamma^2\theta^2/2+\ldots) \quad . \quad . \quad . \quad . \quad . \quad (64)$$

and the corresponding joint function is

$$\begin{aligned}\ln\Psi(\theta_1, \theta_2\,;\ t)=\lambda\alpha t\Big[&\alpha_1\theta_1+\alpha_2\theta_2+\frac{\theta_1^2}{2}(\alpha_1\alpha_2+\alpha_1^2\Gamma^2)\\ &+\frac{\theta_2^2}{2}(\alpha_1\alpha_2+\alpha_2^2\Gamma^2)+\alpha_1\alpha_2\theta_1\theta_2(\Gamma^2-1)+\ldots\Big]. \quad . \quad . \quad (65)\end{aligned}$$

Thus, if we confine our attention to events of group 1,

$$\Gamma_1^2=\alpha_1\Gamma^2+\alpha_2\,; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

and we may deduce the relation

$$\frac{E(n_1n_2)-E(n_1)E(n_2)}{\sqrt{(E(n_1)E(n_2))}}=\sqrt{(\alpha_1\alpha_2)}(\Gamma^2-1). \quad . \quad . \quad . \quad . \quad . \quad (67)$$

These latter results are standard in the theory of partition noise.

It has been pointed out to the author by Professor Pryce that all the above results can be deduced quite generally by an elementary process of averaging, since the partitioning process is independent of the correlating mechanism. In fact this is effectively the method used by Bakker (1938) to deduce the formulæ (66) and (67)*.

We now consider the interval distribution of partitioned events. We first observe that if the original series consists of events of class r, the partitioned series does likewise, since intervals separated by r intervening intervals remain uncorrelated. For events of class 1 the moment generating function of the partitioned series can readily be written down. The problem is then to determine the distribution of $\xi_1+\xi_2+\ldots\xi_r$, where the ξ's are independently distributed with probability $g(\xi)d\xi$, and r has distribution $\alpha_1\alpha_2^{r-1}$. This is a problem arising in many different fields (Kendall 1949) and the solution has been given independently by a number of writers (*e.g.* Woodward 1948). The moment generating function of the moment r's is $\alpha_1/(e^{-\theta}-\alpha_2)$, and that of the ξ's is $G(-\theta)$. Hence the generating function $G^{(1)}(-\theta)$ of the partitioned series is given by

$$G^{(1)}(-\theta)=\alpha_1/[G(-\theta)]^{-1}-\alpha_2. \qquad (68)$$

The first two moments can easily be determined from (68)

$$\left.\begin{aligned} g_1^{(1)}&=g_1/\alpha_1,\\ g_2^{(1)}&=(g_2-g_1^2)/\alpha_1-g_1^2\alpha_2/\alpha_1^2,\end{aligned}\right\} \qquad (69)$$

and we verify from (19) that (66) is satisfied.

For events of class r, although we cannot write down the complete generating function, we can determine the first and second moments. Thus

$$g_1^{(1)}=\underset{r}{\mathrm{E}}\underset{\xi}{\mathrm{E}}(\xi_1+\xi_2+\ldots\xi_r)=\underset{r}{\mathrm{E}}(rg_1)=g_1/\alpha_1. \qquad (70)$$

$$g_{11}^{(1)}=\underset{r}{\mathrm{E}}\underset{\xi}{\mathrm{E}}(\xi_1+\xi_2+\ldots\xi_r)^2=\underset{r}{\mathrm{E}}[rg_{11}+2(r-1)g_{12}+2(r-2)g_{13}+\ldots]$$

$$=\frac{1}{\alpha_1}[g_{11}+2\alpha_2g_{12}+2\alpha_2^2g_{13}+\ldots]. \qquad (71)$$

The correlations can be derived similarly, although the detailed evaluation of the various terms is rather cumbersome. We obtain

$$g_{1s}^{(1)}=\alpha_1^{s-2}\left[g_{1s}+s\alpha_2g_{1(s+1)}+\frac{s(s+1)}{2!}\alpha_2^2g_{1(s+2)}+\ldots(s\geqslant2)\right] \qquad (72)$$

the coefficients of the various terms being those in the expansion of $(1-x)^{-s}$.

---

* I am indebted to Dr. D. K. C. MacDonald for drawing my attention to this paper.

The series (71) (72) are not very convenient for manipulation, since the g_{1s} remain finite for large s. It is more convenient to take away the mean values, and put the results in the form

$$g_{11}^{(1)}-g_1^{(1)2}=\frac{\alpha_2}{\alpha_1^2}g_1^2+\frac{1}{\alpha_1}\left[(g_{11}-g_1^2)+2\alpha_2(g_{12}-g_1^2)+2\alpha_2^2(g_{13}-g_1^2)+\ \dots\right]. \quad (73)$$

$$g_{1s}^{(1)}-g_1^{(1)2}=\alpha_1^{s-2}\left[(g_{1s}-g_1^2)+s\alpha_2(g_{1(s+1)}-g_1^2)+\frac{s(s+1)}{2\,!}(g_{1(s+2)}-g_1^2)+\dots\right].$$

We can then easily verify, on substituting in (34), that the result (66) is obtained for the partitioned series, and this provides a useful check on the internal consistency of the formulæ derived.

ACKNOWLEDGMENT.

The author is indebted to Professor M. H. L. Pryce and Dr. D. K. C. MacDonald for helpful and stimulating discussion, and to the University of Oxford for the award of an I. C. I. Fellowship.

REFERENCES.

BAKKER, C. J., 1938, *Physica*, **5**, 581.
BATEMAN, H., 1910, *Phil. Mag.*, **20**, 698
DOMB, C., 1948, *Proc. Camb. Phil. Soc.*, **44**, 355.
FELLER, W., 1948, *Studies and Essays presented to R. Courant.* Interscience publishers, New York, p. 105.
KENDALL, D. G., 1949, *J. Royal Statist. Soc.*, B, **11**, 230.
NORTH, D. O., 1940, *R. C. A. Review*, **4**, 441.
OWEN, A. R. G., *Proc Roy. Soc.*, B, **136**, (1949) 67.
WHITTAKER, J. M., 1937, *Proc. Camb. Phil. Soc.*, **33**, 451; 1938, *Ibid.*, **34**, 158;
WOODWARD, P. M. 1948, *Proc. Comb. Phil. Soc.*, **44**, 404.

XC. *The Di-Iso-Electronic Row and the Effective Charge in the L Isorows.*

By ANTHONY FELIX RICHTER,
Caroline University, Prague*.

[Received June 16, 1950.]

SUMMARY.

The author's equation for the di-iso-electronic row enables us to compute by extrapolation the electron affinity of hydrogen, $E_H=0{\cdot}74$ eV. The author's equation is nearly identical with that of Hylleraas, but differs in an important term from that of Eucken. A relation is given between screening number and effective charge on one hand, and nuclear charge number Z and number of electrons in the isorow, *i*, on the other. Numerical values of the constants required to calculate screening numbers S_i in the K, L and M—*s* isorows are given in Table I. It appears that with increasing Z, S_i approaches half the coefficient of the linear member of equation (10), *i. e.*, $S_i \sim \frac{1}{2}(2i-1-\bar{a}')$; when $Z \geqslant 20$, the screening constant for the di-iso-electronic row, $S_{II}=0{\cdot}631$, and that for the tri-iso-row, $S_{III}=1{\cdot}63$ and that for the ennea-isorow $S_{IX}=6{\cdot}01$, or $S'_{IX}=6{\cdot}40$ respectively.

IN a previous paper on the K and L electrons (Richter 1948), the ionization potentials of iso-electronic atomic systems were expressed in a quadratic function of the variable $Z+1-i$, where Z is the charge number of the nucleus and *i* the number of electrons surrounding it; the coefficients in the expression are functions of *i* only. The present work relates this method to the usual expression in terms of screening constants.

From the known experimental data we have derived the following expression for the ionization potentials J in a given iso-electronic row

$$J_i=\tfrac{1}{2}B_i(Z+1-i)^2+(A_i-\tfrac{1}{2}B_i)(Z+1-i)+D_i=(Z-S_i)^2R/n^2, \qquad (1)$$

where B_i, A_i and D_i represent the quantum, row and correction constants respectively of the given row. Values of these constants in volts for the K and L iso-electronic rows are given in Table III. of the above-mentioned publication. Since B_i is very nearly the same for both K rows, and again for all L rows, we may take average values B'_K, B'_L, and the reduced constants corresponding to this choice are given in the same table; these reduced constants are distinguished by a dash attached to the symbol. Usually the values of the individual and reduced constants differ little,

e. g.
$$B_{II}=27{\cdot}07, \quad A_{II}=23{\cdot}72, \quad D_{II}=0{\cdot}74,$$
$$B'_K=B'_{II}=27{\cdot}11, \quad A'_{II}=23{\cdot}57, \quad D'_{II}=1{\cdot}02.$$

---

* Communicated by the Author.

The right-hand member of equations (1), where R denotes Rydberg's constant and n the principal quantum number, serves to evaluate the screening number S_i, or the effective charge. On inserting the individual constants for $i=2$ in equation (1) we obtain the equation for the di-iso-electronic row :

$$J_{II}=13{\cdot}535(Z-1)^2+10{\cdot}17(Z-1)+0{\cdot}74=R(Z-S_{II})^2. \quad . \quad . \quad (2)$$

It is worth noting that $10{\cdot}17=3R/4$, corresponding to the K$-\alpha$ line of the X-ray spectra, and further that extrapolation to $Z=1$ gives $E_H=0{\cdot}74$ eV. for the electron affinity of hydrogen. This value is identical with that determined by C. A. Hylleraas (Eucken 1938).

The difference between the di-isoelectronic formula and the quadratic hydrogen term is directly connected with the deviation of the screening number S_i from the value corresponding to "complete" screening, namely i. The screening number for the di-isoelectronic row derived from equation (2) is then given by

$$S_{II}=Z-[Z^2-1{\cdot}263Z+0{\cdot}317]^{\frac{1}{2}}=Z-Z^{*\frac{1}{2}}=Z-Z_{II} \quad . \quad . \quad . \quad (3)$$

or, using the reduced constants :

$$S'_{II}=Z-[Z^2-1{\cdot}260Z+0{\cdot}334]^{\frac{1}{2}}=Z-Z^{*\frac{1}{2}}=Z-Z_{II'}. \quad . \quad . \quad (3.2)$$

Then for $Z\geqslant 20$ we have the simple result

$$S_{II}=0{\cdot}631 \text{ and the effective charge } [Z^*_{II}]^{\frac{1}{2}}=Z-0{\cdot}631. \quad . \quad . \quad (4)$$

Two other equations have been elaborated for the di-isoelectronic row, one by C. A. Hyllerass (Eucken 1938) on wave-mechanical grounds :

$$J_{II}=R[Z^2-1{\cdot}250Z+0{\cdot}31488-0{\cdot}01752Z^{-1}+0{\cdot}00548Z^{-2}]=RZ^*, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and the other by Eucken (1938) :

$$J_{II}=R[Z^2-1{\cdot}250Z]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

It is clear that expressions (2) and (4) differ very little even for small Z, Hylleraas' result giving $S_{II}\sim 0{\cdot}625$. The strange-looking symbol $\sqrt{Z^*_{II}}$ is introduced to accord with the notation of Hylleraas and Hartree (Eucken 1938).

The analogous expressions for the tri-iso-electronic row are as follows :

$$J_{III}=3{\cdot}40(Z-2)^2+2{\cdot}54(Z-2)-0{\cdot}33 \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and

$$S_{III}=Z-[Z^2-3{\cdot}254Z+2{\cdot}401]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

It may be mentioned that the value $2{\cdot}54$ in equation (7) is $3R/16$, corresponding to the L$-\beta$ line of the X spectra. As in the previous case, S becomes practically constant for $Z\geqslant 20$,

$$S_{III}\sim 1{\cdot}63 \quad \text{and} \quad \sqrt{Z^*_{III}}=Z-1{\cdot}63=Z_{III}. \quad . \quad . \quad . \quad (8.2)$$

Our picture of the dependence of the screening number S_i on Z and i becomes more intricate (Richter and Krestýnová 1947) whenever the quantum constant of the iso-row in question differs from that of the hydrogen L level, *i. e.* R/2=6·80. The simple picture may nevertheless be retained, substituting the value for the hydrogen L level either for the individual quantum constant of the L isoelectronic set under examination or for the reduced constant B'_L(=6·90). No very significant error is so introduced, the mean error of the reduced quantum constant (σ=0·033) being less than that of any individual quantum constant in the L isorows with the sole exception of the tri-isoelectronic quantum constant with its σ=0·019. To obtain the dependence in a simple algebraic form, we define quantities $\bar{a}_i$, $\bar{d}_i$, $\bar{a}'_i$, $\bar{d}'_i$, as follows :

$$\bar{a}_i=2A_i/B_i,\quad \bar{d}_i=2D_i/B_i,\quad \bar{a}'_i=2A'_i/B'_i,\quad \bar{d}'_i=2D'_i/B'_i. \quad . \quad . \quad (9)$$

Then the relations between Z and i on one hand and S_i or Z^*_i on the other are expressed by these equations :

$$\bar{S}_i=Z-[Z^2-(i-\bar{a}_i+i-1)Z+(i-\bar{a}_i)(i-1)+\bar{d}_i]^{\frac{1}{2}}=Z-Z_i^{*\frac{1}{2}}, \quad . \quad . \quad . \quad (10)$$

$$\bar{S}'_i=Z-[Z^2-(i-\bar{a}'_i+i-1)Z+(i-\bar{a}'_i)(i-1)+\bar{d}'_i]^{\frac{1}{2}}=Z-\sqrt{Z'^*_i}. \quad . \quad . \quad . \quad (10.2)$$

The calculated values of $\bar{a}_i$, $\bar{d}_i$, $\bar{a}'_i$, $\bar{d}'_i$ necessary to compute S_i and Z^*_i in the first twelve iso-electronic rows are given in Table I. On inserting the appropriate values we can write, *e. g.*

$$Z^*_{IX}=Z^2-12{\cdot}029Z+32{\cdot}35. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and $$Z_{IX}\doteq[Z^2-12{\cdot}029+32{\cdot}35]^{\frac{1}{2}}=Z[1-12{\cdot}029Z^{-1}+32{\cdot}35Z^{-2}]^{\frac{1}{2}}.$$

If

$$|-12{\cdot}029Z^{-1}+32{\cdot}35Z^{-2}|=x_{IX},\ \text{then}\ Z_{IX}=Z(1+x_{IX})^{\frac{1}{2}}. \quad . \quad . \quad (11.2)$$

Under the supposition

$$-1<x<1$$

only the first two terms of Taylor's series are sufficient to determine the practically constant value of the screening power in the isorow inquired into, *i. e.* S_{IX}.

$$Z(1+x_{IX})^{\frac{1}{2}}=Z(1+\tfrac{1}{2}x_{IX}-\tfrac{1}{8}x^2_{IX}-\ .\ .\ .\ )=Z[1+\tfrac{1}{2}x_{IX}+R_2(x_{IX})]. \quad . \quad . \quad . \quad (11{\cdot}3)$$

The error caused by neglecting the remainder $R_2(x_{IX})$ of Taylor's series can be easily approximated by the inegality

$$R_2(x)<\frac{1}{8}\frac{|x|^2}{1-|x|}<\frac{1}{2}\frac{|x\ ^2}{1-|x|} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

giving a more and less approximation.

As x depends on Z what we express by x_Z it is evident that

$$|x_{IX,20}|>|x_{IX,40}|>|x_{IX,80}|.$$

As
$$|x_{IX,20}|=\left|\frac{-12\cdot029}{20}+\frac{32\cdot35}{400}\right|=0\cdot68<1,$$

the formula (11·3) can be limited to the first two terms only without committing a significant error in evaluating S_{IX} for $Z\geqslant 20$.

Therefore
$$Z_{IX}=Z(1-6\cdot01Z^{-1}) \text{ and } S_{IX}=6\cdot01. \quad . \quad . \quad . \quad (11.4)$$

The evaluated error accounts to

$$|R_2(x_{IX})|<0\cdot18<0\cdot72. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.2)$$

In the table of O. Krestýnová (4) the values for $Z=20$, $S_{IX,20}=6\cdot15$ and for $Z=90$, $S_{IX,90}=6\cdot04$ are to be found.

TABLE I.

Constants serving to calculate the screening numbers and the effective charges in the K and L isoelectronic rows according to the formula :

$$S_i=Z-[Z^2-(i-\bar{a}_i'+i-1)Z+(i-\bar{a}_i')(i-1)+\bar{d}_i]^{\frac{1}{2}}=Z-Z^{*\frac{1}{2}}.$$

The numbers with three decimals denote the constants of these isorows ($i=2$, 3, 8 and 9) the empirical quantum numbers of which agree with the hydrogen term exactly.

Isorow i	$\bar{a}_i=\frac{2A_i}{B_i}$	$i-\bar{a}_i$	$i-1$	$\bar{d}_i=\frac{2D_i}{B_i}$	$\bar{a}_i'=\frac{2A_i'}{B_i'}$	$i-\bar{a}_i'$	$\bar{d}_i'=\frac{2D_i'}{B_i'}$
K 1	1	0	0		1	0	0
2	1·752	0·248	1	0·055	1·739	0·261	0·075
L 3	1·746	1·254	2	−0·097	1·62	1·38	−0·06
	1·749	1·251		−0·167			
4	2·40	1·60	3	0·44	2·49	1·51	0·26
	2·55	1·45		0·15			
5	2·54	2·46	4	−0·13	2·67	2·33	−0·35
6	3·33	2·67	5	−0·08	3·28	2·72	0·01
7	3·64	3·36	6	0·73	3·91	3·09	0·33
	3·80	3·20					
8	4·332	3·668	7	−0·435	4·16	3·84	0·17
9	4·971	4·029	8	0·118	4·81	4·19	0·25
10	5·46	4·54	9	0·76	5·45	4·55	0·86
M 11	4·26	6·74	10	−0·79	—	—	
12	5·09	6·91	11	−0·36	—	—	

The table contains in addition to values for the K and L isorows those for the M—s rows.

The quantum constant for the M$-s$ isoelectronic rows, determined from the available data for J, is 3·19, with mean error $\sigma=0{\cdot}04$, and is thus 6 per cent higher than the value of the hydrogen M term. This theoretical difficulty is increased by the fact that the quantum constant of the group M V seems to be clearly different from the constants of the groups M I to M IV; this opinion has been based on the study of the iso-electronic row of the 10th d electron, not only on the M level but also on N (Richter 1947).

From equation (10) it is obvious that the screening power S_i varies with both Z and i, but that for given i it becomes practically constant for all sufficiently large Z; its asymptotic value approaches half the coefficient of the linear term, $2i-1-a'_i$. Since the inter-relations of these isoelectronic equations with the X-ray spectra on one hand, and the screening numbers of A. Sommerfeld, N. Bohr, L. Pauling (Sommerfeld 1931) and J. C. Slater (Slater 1930) on the other, are not entirely elucidated, detailed discussion is reserved for a further paper.

ACKNOWLEDGMENT.

I should like to express my thanks to Miss Anna Berková, cand. med. Prague, and Miss Olga Krestýnová, cand. med. Olomouc, for their careful calculations needed for this and previous publications.

REFERENCES.

EUCKEN, A., 1938, *Lehrbuch der chem. Physik*, Bd. I, 243, 2nd edition.

RICHTER, A. F., 1947, *On the ionization potentials of the M and higher isoelectronic rows expressed as arithmetical progressions of the second degree.* XIth International Congress of pure and applied Chemistry, London. In print; 1948, *Phil. Mag.*, **39**, 303.

RICHTER, A. F., and KRESTÝNOVÁ OLGA, 1947, *Lékařské listy II*, 243.

SLATER, J. C., 1930, *Phys. Rev.*, 2, **36**, 57.

SOMMERFELD, A., 1931, *Atombau und Spektrallinien*, Bd. I, 5 umgearbeitete Auflage, Kap. 5, §4 and Kap. 7, §5.

XCI. *The Laminar Boundary Layer on Solids of Revolution.*

By L. G. WHITEHEAD, Ph.D., and G. S. CANETTI, Ph.D.
Aeronautical Department, Queen Mary College,
University of London *.

[Received June 20, 1950.]

§1. INTRODUCTION.

IMPROVED methods for calculating the laminar boundary layer flow for a cylindrical body were described in an earlier paper, Piercy, Whitehead and Tyler (1948). The purpose of the present paper is to show how these methods can be extended to the solution of the laminar boundary layer equations for the axial flow past a solid of revolution. This problem, which has received less attention than the two-dimensional case, has been examined previously by Millikan (1932) and by Tomotika (1935). In both of these papers a momentum integral relationship similar to von Kármán's two-dimensional equation has been employed to determine the boundary layer thickness. Millikan has assumed a quadratic velocity distribution through the boundary layer, which can give only a rough representation of the flow, while Tomotika has achieved somewhat greater accuracy by employing Pohlhausen's quartic velocity profile.

In the present paper the boundary layer equations for the flow past a solid of revolution are derived in a form similar to that employed in Piercy *et al.* (1948) for two-dimensional flow. Solutions of these equations are found for the flow past cones and are shown to be related to the corresponding two-dimensional flows past wedges given by Falkner and Skan (1930) and Hartree (1937). An approximate solution of general applicability is also presented, which is similar to the two-dimensional approximate solution described in section II. of Piercy *et al.* (1948). A velocity distribution involving three terms is used and these are chosen to give accurate velocity profiles in certain standard cases. Satisfactory results are achieved for the whole of the accelerated part of the boundary layer of slender bodies and for the earlier part of the retarded flow. Except for bluff bodies, a more elaborate solution employing four terms for the velocity profile would be required for the accurate determination of the laminar separation point. At practical Reynolds numbers, however, transition to turbulence will usually occur before the solution using three terms ceases to apply.

The approximate solution, which is exact in the vicinity of the stagnation point of a body with a rounded nose, is compared with the solutions for the flow past a cone and satisfactory agreement is found. The boundary layer flow has also been calculated for a sphere with theoretical and experimental pressure distributions and the results are compared with Tomotika's solutions. In these examples the present method is applicable to the whole laminar boundary layer up to the separation point.

---

*Communicated by Prof. N. A. V. Piercy.

As a final example the flow past an ellipsoid with a fineness ratio of 3 is considered. The boundary layer flow corresponding to the experimental pressure distribution is calculated and a satisfactory solution is obtained over the forward 80 per cent of the laminar layer. The method fails, however, to give a solution close to the laminar separation point.

§2. NOTATION.

The notation used is similar to that employed in the two-dimensional solutions described in Piercy *et al.* (1948) but due to the different nature of the problem the corresponding functions are defined differently in some cases.

A, B, C.	functions of ξ appearing in the series and approximate solutions of the boundary layer equations.
h'	irrotational flow velocity.
h	velocity ratio h'/U.
$h_1\ h_2\ h_3$	functions defining the curvilinear coordinate system.
K	a parameter defining the approximate velocity profile.
l	a representative length of the solid of revolution.
n	degree of the Legendre function $\mathrm{P}_n(\mu)$.
p	static pressure.
q, w	component velocity ratios.
r, θ_1	spherical polar coordinates.
R	Reynolds number.
U	velocity of the undisturbed flow.
V	velocity vector.
x, y	cylindrical coordinates, x measured along the axis of symmetry and y perpendicular to it.
α'	velocity potential of flow with axial symmetry.
β'	Stokes' stream function.
α, β	$\alpha'/\mathrm{U}l$ and $\beta'/\mathrm{U}l^2$ respectively.
δ	boundary layer displacement thickness.
θ	boundary layer momentum thickness.
ϕ, ψ	functions of η.
μ	cos θ_1.
ν	kinematic viscosity.
τ	skin friction.
ξ, η	coordinates defined in terms of α and β.
ζ	vorticity.

§3. BOUNDARY LAYER EQUATIONS.

The exact equations for the incompressible viscous flow with axial symmetry past a solid of revolution are first derived in a modified form. An orthogonal system of independent variables α', β' and γ' are employed.

α'=constant defines an equipotential and β'=constant a streamline for the inviscid axisymmetric flow past the body while γ' is constant on a plane passing through the axis of symmetry. h' is the resultant velocity of the potential flow and q' and w' the velocity components of the viscous flow in directions perpendicular to surfaces α=constant and β=constant respectively. The velocity ratios $q=q'/h'$, $w=w'/h'$ and $h=h'/\mathrm{U}$ are employed in the equations and α' and β' are converted into the non-dimensional forms α and β by dividing by the undisturbed velocity U and by a power of a representative length l. Hence

$$\alpha=\alpha'/\mathrm{U}l\,, \qquad \beta=\beta'/\mathrm{U}l^2$$

and the Reynolds number for the flow is defined as

$$\mathrm{R}=\mathrm{U}l/\nu.$$

The vector form of the equation of motion for steady flow is

$$-\mathbf{V}\times\zeta=-\mathrm{grad}\,(p/\rho+\tfrac{1}{2}\mathbf{V}^2)+\nu\,(\mathrm{grad.\,div}\,\mathbf{V}-\mathrm{curl}\,\zeta)\,, \quad . \quad . \quad . \quad (1)$$

where **V** is the velocity vector and ζ the vorticity. The corresponding form of the equation of continuity for incompressible flow is

$$\mathrm{div}\,\mathbf{V}=0\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The transformation of these equations into curvilinear coordinates is readily carried out. General expressions for the components of the various terms in the two equations are given by Milne-Thomson (1937) in terms of three factors h_1, h_2, and h_3. The element of length along a line perpendicular to the surface α=constant is expressed as $h_1\delta\alpha$ and the corresponding values in the two other directions are $h_2\delta\beta$ and $h_3\delta\gamma$ respectively. It follows therefore that for the present system of orthogonal coordinates

$$h_1=1/h, \quad h_2=1/yh \ \text{ and } \ h_3=y,$$

where y is the radial distance from the axis of symmetry divided by the representative length l. Substitution in (1) yields the equations

$$\frac{1}{h}\frac{\partial h}{\partial\alpha}(q^2+\omega^2)+q\frac{\partial q}{\partial\alpha}+y\omega\frac{\partial q}{\partial\beta}+\frac{\omega^2}{y}\frac{\partial y}{\partial x}$$
$$=-\frac{1}{\rho\mathrm{U}^2h^2}\frac{\partial p}{\partial\alpha}+\frac{1}{\mathrm{R}h^2}\left[\frac{\partial}{\partial x}\cdot h^2\left\{\frac{\partial q}{\partial\alpha}+\frac{\partial}{\partial\beta}(y\omega)\right\}-\frac{\partial}{\partial\beta}\cdot y^2h^2\left\{\frac{\partial}{\partial\alpha}\left(\frac{\omega}{y}\right)-\frac{\partial q}{\partial\beta}\right\}\right], \quad . \quad . \quad . \quad (3)$$

and

$$\frac{1}{h}\frac{\partial h}{\partial\beta}(q^2+\omega^2)+q\frac{\partial\omega}{\partial x}+y\omega\frac{\partial\omega}{\partial\beta}-\frac{q\omega}{y}\frac{\partial y}{\partial\alpha}$$
$$=-\frac{y}{\rho\mathrm{U}^2h^2}\frac{\partial p}{\partial\beta}+\frac{1}{\mathrm{R}h^2}\left[\frac{\partial}{\partial\beta}\cdot h^2\left\{\frac{\partial q}{\partial\alpha}+\frac{\partial}{\partial\beta}(y\omega)\right\}+\frac{\partial}{\partial\alpha}\cdot y^2h^2\left\{\frac{\partial}{\partial\alpha}\left(\frac{\omega}{y}\right)-\frac{\partial q}{\partial\beta}\right\}\right]. \quad . \quad . \quad . \quad (4)$$

A third equation does not arise as there is no flow normal to the planes $\gamma=$constant. The equation of continuity takes the form

$$\frac{\partial q}{\partial \alpha}+\frac{\partial(y\omega)}{\partial \beta}=0. \tag{5}$$

Equations (3) and (4) are simplified by applying the usual assumptions appropriate to flow with a thin viscous boundary layer. In addition the change of y through the boundary layer is neglected so that it can be replaced by y_0, its value on the boundary. This latter assumption will introduce small local errors close to the stagnation point on a solid of revolution with a conical nose since in this region y is comparable with the boundary layer thickness. This does not, however, affect the validity of the solution over the major part of the body. With these approximations equation (4) reduces to

$$q^2\,\frac{1}{h}\,\frac{\partial h}{\partial \beta}=-\frac{y_0}{\rho U^2 h^2}\,\frac{\partial p}{\partial \beta},$$

so that the change of pressure through the boundary layer is negligible. Hence the value of $\partial p/\partial\alpha$ in (3) can be calculated from the irrotational flow just outside the boundary layer, giving

$$\frac{\partial p}{\partial x}=-\rho U^2\left(h\,\frac{dh}{d\alpha}\right)_{\beta=0}$$

Equation (3) therefore reduces to

$$q\,\frac{\partial q}{\partial \alpha}+wy_0\,\frac{\partial q}{\partial \beta}=\left(\frac{1}{h}\,\frac{dh}{d\alpha}\right)_{\beta=0}(1-q^2)+\frac{y_0^2}{R}\,\frac{\partial^2 q}{\partial \beta^2}. \tag{6}$$

This equation is to be solved in conjunction with the equation of continuity (5) for the boundary conditions $q=w=0$ on the solid of revolution, which forms part of the streamline $\beta=0$, and $q=1$ for large values of β.

New variables ξ and η are now introduced in order to simplify the boundary conditions. They are defined in terms of α and β by

$$\xi=\frac{\beta}{2y_0}\left[\frac{R}{\alpha-\alpha_0}\right]^{\frac{1}{2}} \quad\text{and}\quad \eta=\alpha-\alpha_0\,, \tag{7}$$

where α_0 is the value of α at the front stagnation point. The part of the streamline $\beta=0$ lying downstream of the stagnation point corresponds to $\xi=0$ so that the boundary conditions are taken as $q=0$ for $\xi=0$ and $q=1$ for large values of ξ. With this change of variables (6) becomes, after substituting from (5) for w

$$\begin{aligned}\frac{\partial^2 q}{\partial \xi^2}&+\left[2+\frac{4\eta}{y_0}\cdot\frac{dy_0}{d\eta}\right]\frac{\partial q}{\partial \xi}\int_0^{\xi} q\,d\xi\\ &=-4\left(\frac{\eta}{h}\cdot\frac{dh}{d\eta}\right)_{\beta=0}(1-q^2)+4\eta\,\frac{\partial q}{\partial \eta}\cdot q-4\,\frac{\partial q}{\partial \xi}\int_0^{\xi}\frac{\partial q}{\partial \eta}\,d\xi,\end{aligned} \tag{8}$$

an equation which differs only from the two-dimensional form by the inclusion in the second term on the left hand side of the additional factor $4\eta/y_0\,(dy_0/d\eta)$. A function ϕ employed in the two-dimensional solution is reintroduced here together with an additional function ψ. They are defined by

$$\phi=2\left[1-\frac{2\eta}{h}\frac{dh}{d\eta}\right]_{\beta=0} \quad \text{and} \quad \psi=2\left[1-\frac{2\eta}{y_0}\frac{dy_0}{d\eta}\right]_{\beta=0}.$$

Both ϕ and ψ are functions of η only and are zero at the stagnation point of a body with a rounded nose. Equation (8) therefore becomes finally

$$\frac{\partial^2 q}{\partial\xi^2}+(4-\psi)\frac{\partial q}{\partial\xi}\cdot\int_0^\xi q\,d\xi=(\phi-2)(1-q^2)+4\eta\frac{\partial q}{\partial\eta}\cdot q-4\frac{\partial q}{\partial\xi}\int_0^\xi\frac{\partial q}{\partial\eta}.d\xi. \quad . \quad (9)$$

§4. Flow Past Cones.

General solutions of the boundary layer equation (9) give a velocity profile which varies along the surface, but when ϕ and ψ are both constant a solution for q is found which is independent of η. It is readily shown that the flow past a cone is of this type. The velocity potential α is expressed generally in terms of Legendre functions by the relation

$$\alpha=r^n\mathrm{P}_n(\mu),$$

where $\mu=\cos\theta_1$ and r and θ_1 are spherical polar coordinates. θ_0 the value of θ_1 on the conical boundary, is found from the condition $\partial\alpha/\partial\theta_1=0$ and hence is given by the solution of the equation

$$\frac{d}{d\mu}[\mathrm{P}_n(\mu)]=0.$$

θ_0 is the external angle between the axis and the wall of the cone, it has the value $\pi/2$ for $n=2$ giving the flow up to a stagnation point. When $n=1$, $\theta_0=\pi$ and a cone of vanishing thickness results. Values of θ_0 for intermediate values of n, between these two limits, are plotted against π/n in fig. 1. On the conical boundary, η and y_0 are given therefore by

$$\eta=r^n\mathrm{P}_n(\cos\theta_0) \quad \text{and} \quad y=r\sin\theta_0.$$

Hence

$$\phi=-\psi=4/n-2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

With these values of ϕ and ψ substituted in (9), the equation for q, as a function of ξ only, takes the form

$$\frac{\partial^2 q}{\partial\xi^2}+\left(2+\frac{4}{n}\right)\frac{\partial q}{\partial\xi}\int_0^\xi q\,d\xi=4\left(1-\frac{1}{n}\right)(q^2-1). \quad . \quad . \quad . \quad (11)$$

This equation is transformed by the further substitution $\xi_1=\xi(2+4/n)^{-\frac{1}{2}}$ into

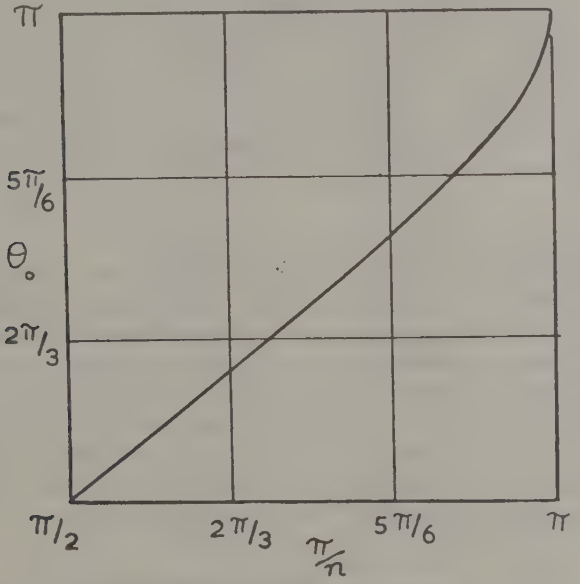
$$\frac{\partial^2 q}{\partial\xi_1^2}+\frac{\partial q}{\partial\xi_1}\int_0^{\xi_1} q\,d\xi_1=\frac{2(n-1)}{n+2}(q^2-1)$$

and if $z'=q$ and $\lambda=2(n-1)/n+2$ it becomes identical with the equation

$$z'''+z''z+\lambda(1-z'^2)=0\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

which has been solved by Hartree for the corresponding two-dimensional problem of the flow past a wedge. Hartree gives accurate numerical values for z for a series of values of the parameter λ. The flow up to the stagnation point on a plane wall ($n=2$) corresponds to $\lambda=\frac{1}{2}$ so that the velocity profile will differ appreciably from that near a two-dimensional stagnation point for which $\lambda=1$. The solution is not valid for values of n close to unity for the boundary layer thickness on the slender cones which occur in this range is comparable with y_0.

Fig. 1.



Variation of cone angle with n.

§5. Series Solution for a Body with a Rounded Nose.

A series solution starting from the front stagnation point can be constructed in a manner similar to that described in Piercy *et al.* (1948) but, due to the presence of the additional function ψ, it will involve, for equal accuracy, the use of a larger number of terms. An expansion for q is assumed in the form

$$q=A+B\phi+C\phi^2+\ldots\ldots, \qquad (13)$$

and each function B, C, etc. is further expanded into series involving functions of ξ multiplied by powers and products of the coefficients in the expansions

$$\eta\frac{d\phi}{d\eta}=a\phi+b\phi^2+c\phi^3+\ldots\ldots,$$

and

$$\psi=a_1\phi+b_1\phi^2+c_1\phi^3.$$

Only the function A and the term in B which is independent of these coefficients have been examined as they are required in the approximate solution given later. They satisfy the equations

$$A''+4A'\int_0^\xi A d\xi-2(A^2-1)=0,$$

and

$$B_0''+4A'\int_0^\xi B_0 d\xi+4B_0'\int_0^\xi A d\xi-4AB_0=1-A^2.$$

The boundary conditions for A are $A=0$ for $\xi=0$ and $A=1$ for ξ large. The equation for A is identical with (11) if n is put equal to 2 so that the solution is given in Hartree's paper. For B_0 the boundary conditions are $B_0=0$ at $\xi=0$ and $B_0=0$ for ξ large. The equation for B_0 has been solved by a method of successive approximations and its values together with A are given in Table I. The second derivatives for all the further functions vanish at $\xi=0$ but otherwise the boundary conditions are identical with those of B_0.

§ 6. THE APPROXIMATE SOLUTION.

An approximate solution which can be employed over the greater part of a laminar boundary layer is next considered. An expression for q is assumed which contains three terms. The first two of these are the initial terms of the series solution and a third term, multiplied by a variable parameter, is used to represent approximately the remaining terms of the series. This type of velocity distribution has given good results for two-dimensional flow and is justified by the fact that the third term replaces terms which have the same boundary conditions and vary in a similar manner throughout the range. Therefore q is expressed as

$$q=A+B_0\phi+K.f(\xi), \quad \ldots \ldots \ldots \quad (14)$$

where the parameter K is a function of η only, and is determined from a momentum integral relationship. A more elaborate solution using a fourth term and two integral relationships would give more accurate results in the retarded region, as is shown in Piercy *et al.* (1948) but the single term used here is adequate for accelerated flow. K is chosen arbitrarily as unity for the flow past a cylindrical tube with its axis lying along the stream. For this case equation (10) reduces to the two-dimensional form for the flat plate since, as y_0 and h are constant, $\phi=\psi=2$. Thus the velocity profile is given by the well known Blasius solution for zero pressure gradient. $f(\xi)$ is chosen therefore so that q takes this form for $\phi=2$ and $K=1$, its values are given in Table I.

The appropriate form of the momentum integral relationship is obtained by integrating (9) through the boundary layer, leading to

$$\left(\frac{\partial q}{\partial \xi}\right)_{\xi=0}=(4-\psi)\int_0^\infty (q-q^2)\,d\xi+(2-\phi)\int_0^\infty (1-q^2)\,d\xi+4\eta\frac{d}{d\eta}\int_0^\infty (q-q^2)\,d\xi. \quad \ldots \quad (15)$$

This equation may be expressed in terms of the two functions

$$\bar{\theta}=\int_0^\infty (q-q^2)\,d\xi \quad \text{and} \quad \bar{\delta}=\int_0^\infty (1-q)\,d\xi,$$

which are related to the momentum thickness θ and the displacement thickness δ of the boundary layer respectively, by

$$\frac{\theta}{l}=\frac{2}{h}\cdot\left(\frac{\eta}{\mathrm{R}}\right)^{\frac{1}{2}}\bar{\theta} \quad \text{and} \quad \frac{\delta}{l}=\frac{2}{h}\left(\frac{\eta}{\mathrm{R}}\right)^{\frac{1}{2}}\bar{\delta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

TABLE I.

ξ	A	$-\mathrm{B}_0$	$f(\xi)$
0·0	0·0	0·0	0·0
0·2	0·3311	0·0510	0·0963
0·4	0·5834	0·0674	0·1839
0·6	0·7615	0·0605	0·2467
0·8	0·8760	0·0432	0·2728
1·0	0·9421	0·0257	0·2609
1·2	0·9760	0·0130	0·2210
1·4	0·9913	0·0056	0·1686
1·6	0·9974	0·0020	0·1173
1·8	0·9993	0·0006	0·0748
2·0	0·9999	0·0002	0·0440
2·2	1·0000	0·0000	0·0241
2·4	.	.	0·0122
2·6	.	.	0·0058
2·8	.	.	0·0025
3·0	.	.	0·0010
3·2	1·0000	0·0000	0·0004

Equation (15) therefore takes the form

$$\left(\frac{\partial q}{\partial \xi}\right)_{\xi=0}=4\eta\,\frac{d\bar{\theta}}{d\eta}+(6-\phi-\psi)\bar{\theta}+(2-\phi)\bar{\delta}, \quad . \quad . \quad . \quad . \quad . \quad (17)$$

and the numerical values of $\bar{\theta}$, $\bar{\delta}$ and $(\partial q/\partial \xi)_{\xi=0}$ appropriate to the chosen velocity distribution are

$$\bar{\theta}=0{\cdot}1752+0{\cdot}0192\phi-0{\cdot}0027\phi^2+0{\cdot}2425\mathrm{K},$$
$$-0{\cdot}0219\phi\mathrm{K}-0{\cdot}0692\mathrm{K}^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

$$\bar{\delta}=0{\cdot}4022+0{\cdot}0550\phi+0{\cdot}3481\mathrm{K}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$(\partial q/\partial \xi)_{\xi=0}=1{\cdot}8554-0{\cdot}3523\phi-0{\cdot}4868\mathrm{K}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The non-dimensional skin friction coefficient $\tau\mathrm{R}^{\frac{1}{2}}/\frac{1}{2}\rho\mathrm{U}^2$ is expressed in terms of $(\partial q/\partial \xi)_{\xi=0}$ by the relation

$$\frac{\tau\mathrm{R}^{\frac{1}{2}}}{\frac{1}{2}\rho\mathrm{U}^2}=\frac{h^2}{\eta^{\frac{1}{2}}}\left(\frac{\partial q}{\partial \xi}\right)_{\xi=0}$$

A differential equation for K is obtained by substituting these numerical values in (17) and it is solved for each example by a step by step method of integration. The equation is

$$\begin{aligned}4\eta\, d\mathrm{K}/d\eta\,(0{\cdot}2425-0{\cdot}0219\phi\;-0{\cdot}1384\mathrm{K})\\ =0{\cdot}0904\phi^2-2{\cdot}6379\mathrm{K}+0{\cdot}4150\mathrm{K}^2+0{\cdot}7221\phi\mathrm{K}\\ -4\eta\, d\phi/d\eta\;(0{\cdot}0192\;-0{\cdot}0054\phi\;-0{\cdot}0219\mathrm{K})\\ +\psi\,(0{\cdot}1752\;+0{\cdot}0192\phi\;+0{\cdot}2425\mathrm{K})\\ -(\phi+\psi)\,(0{\cdot}0027\phi^2+0{\cdot}0219\phi\mathrm{K}+0{\cdot}0692\mathrm{K}^2).\end{aligned}$$

A first check on the accuracy of the approximate solution is made by a comparison with the accurate solutions for the flow past a cone found in §4. For these examples, as ϕ and ψ are constant, K is also constant and equation (21) reduces to the simple quadratic for K

$$0{\cdot}4150\mathrm{K}^2+\left(\frac{1{\cdot}9184}{n}-3{\cdot}5971\right)\mathrm{K}+\frac{1{\cdot}1392}{n^2}-\frac{1{\cdot}8400}{n}+0{\cdot}6352=0.$$

Values of K for a range of n between 1 and 2 are given in Table II. and the approximate values of $(\partial q/\partial\xi)_{\xi=0}$ are compared with values found by Tomotika's method and with the accurate values obtained by Hartree. The accuracy for the present approximate solution is seen to be superior to that of Tomotika's solution, the maximum error for the range considered being 0·6 per cent compared with 1·8 per cent.

TABLE II.

λ	n	K	$\left(\frac{\partial q}{\partial\xi}\right)_{\xi=0}$ Present method	$\left(\frac{\partial q}{\partial\xi}\right)_{\xi=0}$ Tomotika	$\left(\frac{\partial q}{\partial\xi}\right)_{\xi=0}$ Hartree
0·1	1·159	−0·0529	1·369	1·380	1·371
0·2	1·333	−0·0477	1·526	1·525	1·536
0·3	1·530	−0·0343	1·655	1·639	1·665
0·4	1·750	−0·0176	1·763	1·736	1·768
0·5	2·000	0·0000	1·855	1·832	1·855

§7. THE LAMINAR BOUNDARY LAYER ON A SPHERE.

The boundary layer flow on a sphere has been calculated previously by Tomotika for the potential flow pressure distribution and by Tomotika and Imai (1938) for an experimental pressure distribution measured by Fage (1936). Solutions for these examples have been calculated by the present method and are compared with these earlier results.

For the irrotational flow past the sphere the velocity near the boundary is given by $h=\frac{3}{2}\sin\theta_1$ and, if the representative length is taken as the diameter, $\eta=\frac{3}{4}(1-\cos\theta_1)$ and $y_0=\frac{1}{2}\sin\theta_1$. The corresponding values of ϕ, ψ and $\eta\,(d\phi/d\eta)$ are therefore

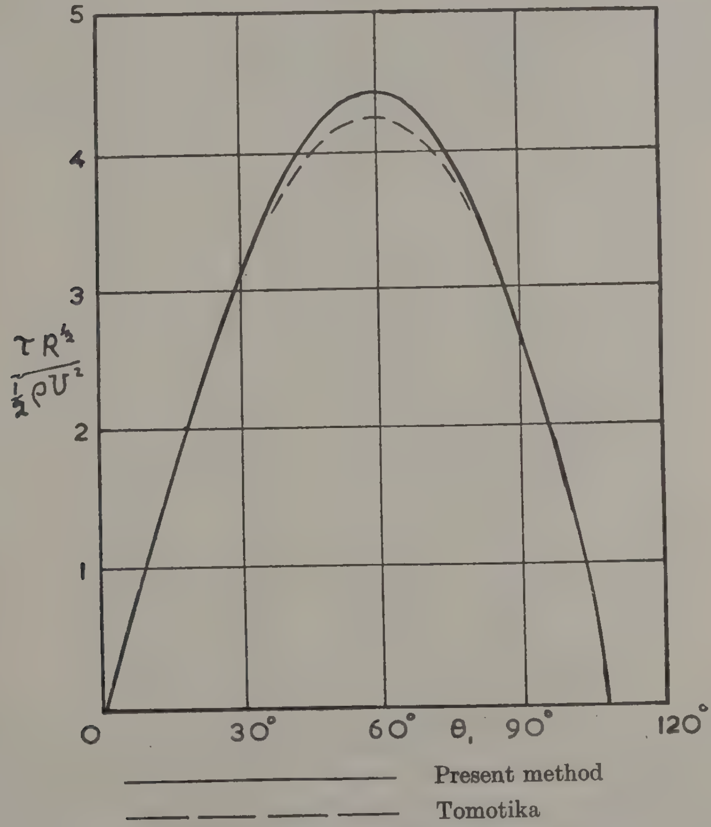
$$\phi=\psi=\frac{2(1-\cos\theta_1)}{1+\cos\theta_1},\qquad \eta\frac{d\phi}{d\eta}=\frac{4(1-\cos\theta_1)}{(1+\cos\theta_1)^2},$$

and it is convenient to change the variable η in equation (21) to θ_1 for the step by step integration. A series solution for small values of θ_1 shows that only even powers arise and the starting value of K is given by

$$K=0\cdot01365\,\theta_1^2+\ldots\ldots$$

The skin friction coefficient $\tau R^{\frac{1}{2}}/\frac{1}{2}\rho U^2$ is plotted as the continuous line in fig. 2 and the dotted line shows Tomotika's solution. The separation point is found to be 107° from the front stagnation point compared with 108° for his solution.

Fig. 2.



Skin friction on sphere, theoretical pressure distribution.

Tomotika and Imai have expressed the experimental velocity distribution on a sphere measured by Fage for an entirely laminar boundary layer as

$$h=1\cdot5\,\theta_1-0\cdot43707\,\theta_1^3+0\cdot148097\,\theta_1^5-0\cdot042329\,\theta_1^7,$$

and this form is retained for the calculation of ϕ and ψ in the present solution, though it differs slightly from the measured values in the immediate vicinity of the separation point. ϕ, ψ and η $(d\phi/d\eta)$ are calculated from the relations

$$\eta=\int h d\theta_1,$$

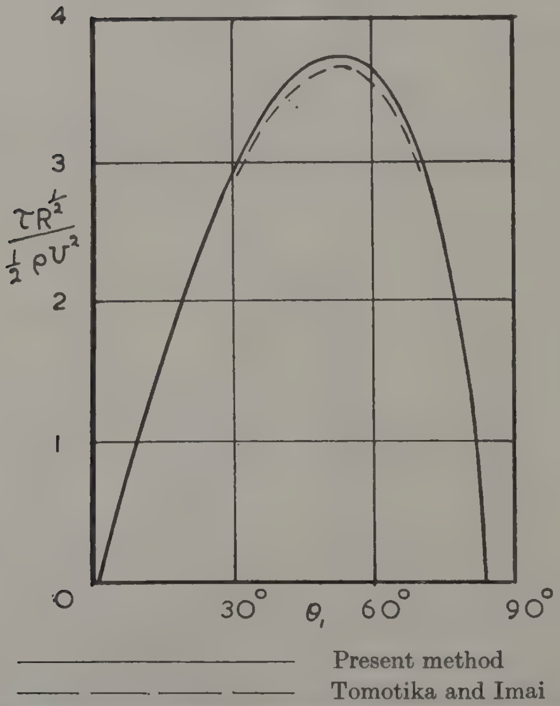
$$\phi=2-\frac{4\eta}{h^2}\frac{dh}{d\theta_1},$$

$$\psi=2-\frac{4\eta}{h}\,.\cot\theta_1,$$

and

$$\eta\frac{d\phi}{d\eta}=\frac{4\eta}{h^2}\left\{\frac{2\eta}{h^2}\left(\frac{dh}{d\theta_1}\right)^2-\left(\frac{dh}{d\theta_1}\right)-\frac{\eta}{h}\frac{d^2h}{d\theta_1^2}\right\}.$$

Fig. 3.



Skin friction on sphere, experimental pressure distribution.

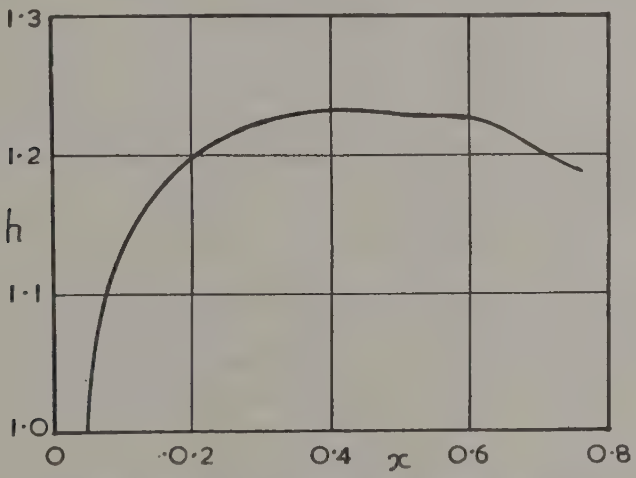
The inital values for the step by step integration of (21) are given by $K=-0{\cdot}00035\ \theta_1^2$ in this case. The skin friction coefficient for the experimental pressure distribution is plotted as the continuous line in fig. 3 and the dotted line gives Tomotika and Imai's solution. Somewhat higher values for the skin friction are given by the present method for both pressure distributions. Tomotika's method is a development of

Pohlhausen's two-dimensional solution which has been compared by Bairstow (1939) with more accurate solutions for the flow past a circular cylinder. In this two-dimensional problem it is found that Pohlhausen's method underestimates the friction and it is likely that Tomotika's solution for the sphere will involve similar errors. The separation point calculated from the experimental pressure distribution is 83·3° from the front stagnation point compared with 84° for Tomotika's solution and the experimental value of 83° found by Fage.

§8. FLOW PAST AN ELLIPSOID.

The boundary layer flow past an ovary ellipsoid of fineness ratio 3 is next considered. This example involves much smaller pressure gradients than occur on the sphere over the major part of the boundary layer and in this respect it is more representative of a practical streamline shape. The experimental pressure distribution has been measured at a Reynolds' number of $2{\cdot}5\times10^5$ based on the major axis, which is 1 ft. long. The ellipsoid was mounted in a 4 ft. square wind tunnel and supported by a tube 1 in. diameter which formed a rearward extension of the ellipsoid and which was faired into it. All pressure connections were carried through this tube and the pressures were recorded after carefully aligning the axis of the ellipsoid along the stream. In fig. 4, values of h calculated from these pressure readings are plotted against x, the distance from the nose divided by the major axis. The readings are plotted as far as the laminar separation point which is located at $x=0{\cdot}76$. The skin friction calculated by the present method is shown in fig. 5. The solution has not been carried beyond $x=0{\cdot}65$ as its accuracy is considered doubtful further downstream. A difficulty arises in the retarded part of the flow when K approaches a value of about 1·4 as the coefficient of $dK/d\eta$ decreases rapidly

Fig. 4.



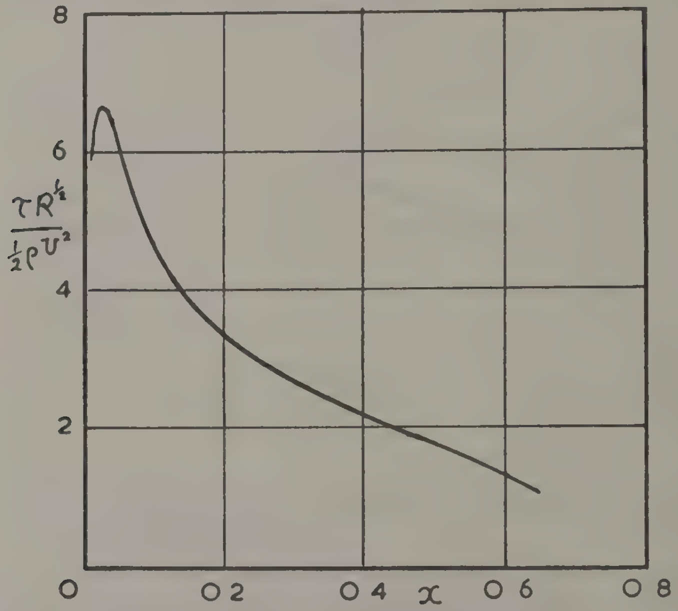
Velocity outside the boundary layer of the ellipsoid.

and becomes negative for larger values of K. For this reason the method fails to give a solution near the laminar separation point in the present example.

§ 9. CONCLUSION.

Solutions for the laminar boundary layer flow on a solid of revolution are described which employ methods previously applied successfully in two-dimensional flow. An approximate solution is obtained which is superior in accuracy to earlier methods for accelerated boundary layer flow. The approximate solution is compared with accurate solutions for the flow past cones and satisfactory agreement is found. Comparisons with Tomotika's solutions for the flow past a sphere show that the skin friction calculated by the present method is larger over most of the laminar layer.

Fig. 5.



Skin friction for flow past ellipsoid.

REFERENCES

L. BAIRSTOW, 1939, *Applied Aerodynamics* (2nd Edition) (London : Longmans, Green and Co.) pp., 524, 543.

A. FAGE, 1936, *A.R.C.R. & M.*, No. 1766.

V. M. FALKNER and S. W. SKAN 1930, *A.R.C.R. & M.*, No. 1314.

D. R. HARTREE 1937 *Proc. Camb. Phil. Soc.*, **33,** 223-239.

C. B. MILLIKAN 1932, *Trans. Amer. Soc. Mech. Eng., Appl. Mech. Sec.*, **54,** 29.

L. MILNE-THOMSON, 1937, *Theoretical Hydrodynamics* (1st Edition) (London : Macmillan) pp. 55-58.

PIERCY, N.A. V., WHITEHEAD, L. G., and TYLER, R. A., 1948, *Aircraft Engineering,* **20,** 354-359.

TOMOTIKA, S., 1935, *A.R.C.R. & M.*, No. 1678.

TOMOTIKA, S, and IMAI, I, 1938, *Proc. Phys.-Math. Soc. Japan*, 3rd Ser. **20,** 287-303.

XCII. *A Theory of Radar Information.*

By P. M. Woodward, B.A. and I. L. Davies, M.A.,
T.R.E., Ministry of Supply *.

[Received April 13, revised July 18, 1950.]

Summary.

The theoretical accuracy and certainty with which range may be determined by radar is obtained quantitatively by applying the principle of inverse probability. In agreement with experience, the theory yields two fundamental criteria for satisfactory radar reception. First, the total received signal energy must always be larger than the effective noise power per unit bandwidth. Secondly, there is a more stringent threshold of unambiguous reception which depends not only on the quantity of received energy but also on the bandwidth of the transmitted waveform, for as this is increased the minimum energy required for the certain detection of an echo also increases. The quantity of information obtainable at the receiver, as measured by reduction of entropy, is evaluated for comparison with that given by Shannon's general theory of communication (Shannon 1948). It is found that as the time of observation and hence the received energy increases, information is initially obtained at a roughly uniform rate which is not far removed from the absolute limit for an ideal communication system, but that after crossing the threshold of unambiguous reception, additional received energy contributes little further information.

Since only stationary echoing objects are considered, phase-coherent integration is implied in the theory and the problem of post-detection integration is not discussed. Apart from obviously unsuitable waveforms, the theory is applicable to any type of modulation.

§1. Introduction.

It has long been recognized that the amount of information which can be obtained from any physical observation is fundamentally limited by random fluctuations. In communication problems, for example, the physicist or engineer always strives to work with signals which are large compared with the background of noise. Superficially, one might be tempted to conclude that information can be communicated at a satisfactory rate only when the signal power P is greater than the noise power N in the system. There is, however, no such abrupt limitation, for Shannon (1948, 1949) has shown that the limiting rate at which information can be communicated, in the presence of a uniform spectrum of Gaussian noise, is given by

$$C=W\log_2\left(1+\frac{P}{N}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

---

* Communicated by the Authors.

binary choices per unit time, when no frequencies higher than W are utilized. Nevertheless, many practical systems do exhibit a threshold of intelligibility when P is equal to N, but this is only due to the manner of coding the information. It is not revealed by equation (1) because this assumes ideal coding, which might be somewhat complicated if P were less than N. A threshold is especially pronounced when the information is encoded by a non-topological increase in its natural number of dimensions (Shannon 1949).

The simplest information sought from a radar system is the range of an echoing object, presumed stationary in this paper. Such information has only a single dimension, but it is coded (perforce) in a special manner which increases its dimensionality. A repetitive waveform, not necessarily a sequence of pulses, is transmitted, echoed, and received. Observation is assumed to last only for a finite time—any integral number of repetition periods—during which only a finite quantity of signal energy is available at the receiver: the observer compares the transmitted and received waveforms and tries to determine the time-delay between them. The presence of noise in the received waveform prevents him doing so with unlimited precision. Both transmitted and received waveforms occupy a substantial area of the time-frequency information diagram as defined by Gabor (1946), and therefore encompass a large number of elementary cells. In Shannon's terminology, their dimensionality is high. But since the shape of the transmitted waveform is fixed at the outset and contains no information at all, and the received waveform is only a delayed copy, it is not obvious that this time-frequency area can be utilized to the same advantage as in an ordinary communication system. One of the objects of this paper is to show that radar can be fairly efficient in this fundamental sense. The primary purpose of the work, however, is to determine the threshold of intelligibility, which we might expect to be quite marked in view of the nature of the coding. Under conditions of intelligible reception, it is further possible to determine the inherent accuracy with which target range may be ascertained.

The mathematical analysis centres round one important concept, the *a posteriori* probability distribution for the time-delay τ between transmission and reception. This probability distribution describes all that can possibly be known about range after a radar observation has been made, and is formally derived in §3 after a preliminary statement of terminology in §2. The distribution will be found to depend on two significant functions of τ, for which the terms signal function and noise function are adopted. They are, effectively, the signal auto-correlation and signal-noise cross-correlation functions (Wiener 1949). Both of them depend on the known shape of the transmitted waveform and on the known statistical characteristics of the noise, but apart from this, they depend separately on the signal and noise actually received. Their relative importance is discussed in §4, where certain of their properties are derived for subsequent reference. The so-called signal function shows to what extent the accuracy of a range determination is, in the presence of noise inherently limited (§5), but the limitation it describes is merely the "topological uncertainty" of the observation, because it refers to a connected region of uncertainty in the neighbourhood of the true value of the range. The noise function, on the other hand, may produce spurious peaks in the probability distribution, representing "non-topological uncertainty" or ambiguity which may exist as the result of an observation, and which may render the observation valueless from a practical point of view. Such ambiguity arises from the non-topological mapping referred to earlier. The conditions under which ambiguity occurs in a radar system are examined in §6, where a formula for the threshold of intelligibility is derived—fig. 1 illustrates the effect. Finally, in §7, the entropy of the *a posteriori* distribution is obtained: after subtraction from the *a priori* entropy it represents the information content of the observation (fig. 2), and it is interesting to compare this with Shannon's general formula for the information capacity of a communication channel. The main features of the results are briefly discussed in the concluding section.

§ 2. REPRESENTATION OF WAVEFORMS.

The transmitted radio-frequency waveform will be denoted by the complex function $\psi(t)$, and the received waveform by

$$\gamma(t)=\alpha\psi(t-\tau)+\nu(t). \qquad (2)$$

Upon reception, $\psi(t)$ will have suffered a change in amplitude and a delay of τ in its passage to and from the target, while $\nu(t)$ is a complex random function representing added noise. These complex functions are all formed from the physical waveforms by the addition of an imaginary quadrature component in the manner described, for instance, by Gabor (1946). Doing the analysis in terms of such complex waveforms introduces no spurious information, and it is only necessary to remember that the energy which would be dissipated in a unit resistance is no longer the time integral of the square of the waveform, but half the integral of the squared modulus. As a matter of purely mathematical convenience, it will be presumed that the noise energy is confined within a finite band of frequencies. As long as this band is sufficiently wide to include the whole spectrum of ψ, there is no loss of generality, for it is merely equivalent to assuming that $\gamma(t)$ has been filtered so as to reduce the noise while leaving the signal unaffected. Such curtailing of the noise, which is assumed uniformly distributed in frequency, has no bearing on the problem of optimum filtering.

Since γ, ψ, and ν are all high-frequency functions, they may be written in the form

$$\left.\begin{aligned}\gamma&=y(t)\exp(2\pi i f_0 t),\\ \psi&=u(t)\exp(2\pi i f_0 t),\\ \nu&=n(t)\exp(2\pi i f_0 t),\end{aligned}\right\} \qquad (3)$$

where y, u and n are complex low-frequency functions, and f_0 is the carrier frequency defined by

$$f_0 = \frac{1}{2\pi i} \frac{\int \psi^* \frac{d}{dt} \psi \, dt}{\int \psi^* \psi \, dt}. \qquad (4)$$

It may be shown (Gabor 1946) that this definition makes f_0 the centroid of the power spectrum of ψ. When ψ is periodic, it is equivalent to

$$\int |\psi|^2 (f - f_0) \, dt = 0, \qquad (5)$$

where the integral is taken over one period and f denotes the instantaneous frequency, or rate of change of phase, of $\psi(t)$. The above definitions have the merit of applying equally well to any form of modulation, and are essential to the subsequent analysis.

Further, it is convenient to introduce here the vectorial representation of waveforms (Shannon 1949, Oswald 1949). Suppose that $z(t)$ is a complex function of t, whose complex frequency spectrum $s_z(f)$ is zero outside the range of frequencies $(-\mathrm{W}, \mathrm{W})$, as is assumed for the particular functions y, u, and n defined above. It is well known that $z(t)$ may be completely specified by a set of sample values $z(r/2\mathrm{W})$, and may therefore be represented by a complex multi-dimensional vector $\mathbf{z}$ with these components. The Hermitian product of two such vectors $\mathbf{z}$ and $\mathbf{w}$, is

$$\mathbf{z}^\dagger \mathbf{w} = \sum_r z_r^* w_r = 2\mathrm{W} \int z^*(t) w(t) \, dt. \qquad (6)$$

The second equation may be proved by applying Parseval's theorem (Titchmarsh 1939) successively to the Fourier Integrals representing $z(t)$ and $w(t)$ and to the Fourier Series representing $s_z(f)$ and $s_w(f)$ in the interval $(-\mathrm{W}, \mathrm{W})$. It follows immediately from equation (6) that

$$\mathbf{u}^\dagger \mathbf{u} = |\mathbf{u}|^2 = 2\mathrm{W} \int |u(t)|^2 \, dt = 4\mathrm{WE}, \qquad (7)$$

where E denotes the energy of the waveform u, or ψ, within the range of integration. Further, if $\mathbf{u}'$ represents the waveform du/dt, we may define an effective bandwidth β by the equation

$$|\mathbf{u}'|^2 = \beta^2 |\mathbf{u}|^2. \qquad (8)$$

This bandwidth β is, apart from a constant factor, the effective bandwidth as defined by Gabor (1946), who has shown that $\beta^2/4\pi^2$ is the second moment of the power spectrum of u about the origin, or ψ about f_0.

Vector notation is especially convenient when dealing with noise characteristics, and the results quoted in this section will be used in vector and integral form indiscriminately, since each may be transformed into the other by equation (6).

§ 3. BASIC THEORY OF RADAR OBSERVATION.

The central problem is to form, from any particular received waveform and from a knowledge of the transmitted waveform, the probability distribution for the target range, measured in terms of the time-delay τ. Suppose that, on some typical occasion, the actual value of the time-delay is τ_0 and of the coefficient α is α_0; the received waveform is then

$$\gamma(t)=\alpha_0\psi(t-\tau_0)+\nu(t). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

In order to form the probability distribution for τ (and in general α also) when only $\gamma(t)$ and $\psi(t)$ are known, one must first determine the probability density for the observed $\gamma(t)$ on the assumption of any guessed τ and α. This depends solely on the noise which would be necessary to satisfy the assumption. On a hypothesis (τ, α), the noise would in fact have to be

$$\gamma(t)-\alpha\psi(t-\tau),$$

or in terms of the low-frequency functions defined in equation (3),

$$\{y(t)-\alpha u(t-\tau)\exp(-2\pi i f_0\tau)\}\exp(2\pi i f_0 t). \quad . \quad . \quad . \quad . \quad (10)$$

The expression in large brackets may be represented by the vector

$$\mathbf{y}-\alpha\mathbf{u}(\tau)e^{-i\omega\tau}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where $u(t-\tau)$ has become a continuous vector function of τ, and ω stands for $2\pi f_0$. The vector (11) uniquely determines the high frequency noise fluctuation, and assuming Gaussian noise, its multi-dimensional probability density is proportional to

$$\exp\{-|\mathbf{y}-\alpha\mathbf{u}(\tau)e^{-i\omega\tau}|^2/2\mathrm{N}\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Here N is the mean physical noise power, and is the mean square value of both the real and imaginary part of the noise fluctuation at each sampling point. All these values are uncorrelated because the noise is assumed to extend *uniformly* over a high-frequency band of width 2W.

If the *a priori* probability distribution for τ and α is denoted by $p_0(\tau, \alpha)$, then by the principle of inverse probability (Jeffreys 1939) their *a posteriori* distribution is

$$p(\tau, \alpha)\propto p_0(\tau, \alpha)\exp\{-|\mathbf{y}-\alpha\mathbf{u}(\tau)e^{-i\omega\tau}|^2/2\mathrm{N}\}. \quad . \quad . \quad (13)$$

This is the frequency distribution of the true values of τ and α over an ensemble in which the received waveform $\gamma(t)$ is fixed. The *a priori* distribution, on the other hand, is the corresponding frequency over the ensemble of all waveforms received, and must be precisely formulated before one can proceed to analyse the problem further.

Since this paper is concerned with the information content of received signals of given energy, it is assumed that the coefficient α is constant, and hence that its *a priori* distribution is given by

$$p_0(\alpha)=\delta(\alpha-\alpha_0). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Further, it is assumed that the *a priori* distribution of τ is uniform over a fixed interval T, less than or equal to the repetition period of the transmitted waveform. Such an assumption is not only the simplest but also the most non-commital one to make, because it represents a state of least prior knowledge (Jeffreys 1939) between the limits of the distribution. Within the interval T we therefore write

$$p_0(\tau, \alpha)=\frac{1}{\mathrm{T}}\delta(\alpha-\alpha_0). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

It follows from equation (13) that the *a posteriori* distribution for τ is

$$p(\tau)=\int_0^\infty p(\tau, \alpha)\,d\alpha \propto \exp\{-\,|\,\mathbf{y}-\alpha_0\mathbf{u}(\tau)e^{-i\omega\tau}\,|^2/2\mathrm{N}\}, \quad . \quad (16)$$

within the *a priori* interval T, and zero elsewhere.

This expression will now be reduced to a more convenient form. By expanding, and omitting terms independent of τ, we obtain

$$p(\tau)\propto \exp\left[\frac{\alpha_0}{\mathrm{N}}\mathscr{R}\,\mathbf{y}^\dagger\{\mathbf{u}(\tau)e^{-i\omega\tau}\}\right]. \quad . \quad . \quad . \quad . \quad (17)$$

Of the terms which have disappeared, $|\,\mathbf{y}\,|^2$ is dependent on τ_0 but not on τ, whilst the term $|\,\mathbf{u}(\tau)\,|^2$ is proportional to the energy of the signal over the interval of observation, and is independent of τ because it extends over an integral number of periods. Converting now into integral form,

$$p(\tau)\propto \exp\left\{\frac{\alpha_0}{\mathrm{N}_0}\mathscr{R}\int y^*(t)u(t-\tau)e^{-i\omega\tau}\,dt\right\}, \quad . \quad . \quad . \quad (18)$$

where

$$\mathrm{N}=2\mathrm{W}\mathrm{N}_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

and N_0 is the mean noise power per unit bandwidth.

From equations (3) and (9), substitution for y gives

$$p(\tau)\propto \exp\mathscr{R}\left[\frac{1}{\mathrm{N}_0}e^{-i\omega(\tau-\tau_0)}\int\{\alpha_0^2u^*(t-\tau_0)u(t-\tau) + \alpha_0 n^*(t)u(t-\tau)e^{-i\omega\tau_0}\}\,dt\right]. \quad . \quad . \quad . \quad . \quad (20)$$

Since $n(t)$ is a complex random quantity, the constant $\exp(-i\omega\tau_0)$ may be absorbed into it without altering its statistical properties, and is in future omitted. Denoting the phase of the integral by θ, and writing

$$g(\tau)=\frac{\alpha_0^2}{\mathrm{N}_0}\int u^*(t-\tau_0)u(t-\tau)\,dt \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$h(\tau)=\frac{\alpha_0}{\mathrm{N}_0}\int n^*(t)u(t-\tau)\,dt, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

one obtains equation (20) in the form

$$p(\tau)\propto \exp[\cos\{\omega(\tau-\tau_0)-\theta\}\,.\,|\,g(\tau)+h(\tau)\,|]. \quad . \quad . \quad . \quad (23)$$

The cosine is a high-frequency function of τ, while θ and also the modulus are low-frequency functions. This means that the *a posteriori* distribution consists of a closely-packed sequence of peaks under a slowly varying envelope; these peaks arise from the comparison which the observer could make between the carrier phases of transmitted and received waveforms.

The information which this fine structure represents is of no value when there is ambiguity between one peak and another; the realistic distribution for τ, denoted $P(\tau)$ is consequently given by smoothing $p(\tau)$ to remove fine structure. It is important, however, to realize that, although the information which could be obtained by comparing the phase of the transmitted and received carriers will be discarded, phase-coherence from one repetition cycle of the modulation to the next is assumed throughout the interval of observation. In other words, variation of phase yields useful information but absolute magnitude does not. The smoothed distribution is easily obtained by integrating equation (23), with respect to τ, from τ to $\tau+(2\pi/\omega)$, and assuming that the variation of θ and the modulus are negligible over this interval. This is not equivalent to taking the envelope of $p(\tau)$, because the fine structure is not of a sinusoidal nature. Integration gives, in fact,

$$P(\tau)=\lambda I_0(|g+h|), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where I_0 is a modified Bessel function, and λ is a normalizing constant. It should be emphasized that the integrals g and h extend over the whole interval of observation.

The characteristics of $P(\tau)$ depend markedly on the relative importance of g and h, which may be regarded as the contributions of the signal and noise to the probability distribution. For convenience, they are termed the signal and noise function respectively: they are, in effect, the signal auto-correlation and signal-noise cross-correlation functions (Wiener 1949).

§ 4. THE SIGNAL AND NOISE FUNCTIONS.

In this section, some of the properties of $g(\tau)$ and $h(\tau)$ will be formulated, and there will emerge, incidentally, a criterion for the amount of energy which must be received from an echo before satisfactory radar observation is possible.

We should expect $P(\tau)$ to have a peak in the vicinity of the true range τ_0; this will arise from the signal function $g(\tau)$. Elsewhere, $g(\tau)$ will in practice be negligible because, in effect, radar waveforms are specially chosen to ensure it. Near τ_0 one may expand $u(t-\tau)$ by Taylor's theorem, and obtain from equation (21)

$$g(\tau)=\frac{\alpha_0^2}{N_0}\int u^*(t-\tau_0)\{u(t-\tau_0)-(\tau-\tau_0)u'(t-\tau_0) + \tfrac{1}{2}(\tau-\tau_0)^2u''(t-\tau_0)\}\,dt, \quad . \quad . \quad . \quad (25)$$

neglecting higher powers of $\tau-\tau_0$ and denoting du/dt by u'. Writing

$$\rho^2=2E/N_0, \qquad (26)$$

where E is now the energy of the received signal $\alpha_0\psi$, one may obtain from equations (7), (4) and (8) the following useful relations :

$$\frac{\alpha_0^2}{N_0}\int u^*(t-\tau_0)u(t-\tau_0)\,dt=\rho^2, \qquad (27)$$

$$\frac{\alpha_0^2}{N_0}\int u^*(t-\tau_0)u'(t-\tau_0)\,dt=0, \qquad (28)$$

$$\frac{\alpha_0^2}{N_0}\int u^*(t-\tau_0)u''(t-\tau_0)\,dt=-\rho^2\beta^2. \qquad (29)$$

These are definite integrals, extending over an integral number of repetition periods of $u(t)$. The second relation depends on the definition (4) of carrier frequency, while the third is obtained after an integration by parts. Consequently,

$$g(\tau)=\rho^2-\tfrac{1}{2}\rho^2\beta^2(\tau-\tau_0)^2+O(\tau-\tau_0)^3. \qquad (30)$$

It will be seen that the peak in $g(\tau)$ is of height ρ^2 occurring at τ_0, and in this neighbourhood $g(\tau)$ is purely real. This quantity ρ^2 is important because it is the dimensionless energy parameter which constantly recurs in the theory. The expansion of $g(\tau)$ outlined above is required later, when determining the accuracy of range measurement.

It is now necessary to consider the noise function $h(\tau)$ and its influence on $g(\tau)$. Transformation of equation (22) into vector notation gives

$$h(\tau)=\frac{\alpha_0}{N}\mathbf{n}^\dagger\mathbf{u}(\tau). \qquad (31)$$

This Hermitian product is a linear function of the components of $\mathbf{n}$, and the resultant is a stationary random function whose distributions in real and imaginary parts are independent and Gaussian about zero. By using the additive property of mean square deviations, one may readily show that

$$\overline{\{\mathscr{R}h(\tau)\}^2}=\overline{\{\mathscr{I}h(\tau)\}^2}=\alpha_0^2\,|\,\mathbf{u}(\tau)\,|^2/N=\rho^2. \qquad (32)$$

Furthermore, it is well known (Rice 1945) that the amplitude distribution for such a function is given by

$$q(\,|\,h\,|\,)=\frac{1}{\rho^2}|\,h\,|\exp\left(-\,|\,h\,|^2/2\rho^2\right), \qquad (33)$$

and that $|\,h\,|$ has a mean square value $2\rho^2$.

When $E<N_0$, the root mean squared modulus, $\rho\sqrt{2}$, of $h(\tau)$ is greater than the peak value ρ^2 of $g(\tau)$, and the signal and noise functions will not be distinct in $P(\tau)$. A necessary, though insufficient, criterion for satisfactory observation is therefore that the total received energy should be

greater than the noise power per unit bandwidth, and for convenience in the mathematics it will be assumed much greater. The condition

$$\rho^2 \gg 2, \quad \text{or} \quad E \gg N_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

applies, in fact, throughout the remainder of this paper. It is insufficient because it does not of itself ensure that the area under $P(\tau)$ is all near τ_0.

An immediate consequence of this condition is that the peak of the signal function $g(\tau)$ will be large compared with unity, and it can be seen from equation (24) that this feature is accentuated in $P(\tau)$, since asymptotically $I_0(x)$ increases exponentially with x. Thus, for practical radar waveforms, the peak in the normalized probability distribution due to the signal function will be concentrated near τ_0 and in equation (30) the higher powers of $(\tau-\tau_0)$ may be neglected. Except near this peak, the signal function will have a negligible influence on $P(\tau)$ and the noise function (if it is of sufficient magnitude) will alone affect it. The total interval T over which $P(\tau)$ is defined will thus be considered in two parts : one which includes the peak of $P(\tau)$ due to the signal function, where $P(\tau)$ will be denoted by $P_g(\tau)$, and the remainder, of a length nearly equal to T, where $P(\tau)$ will be denoted by $P_h(\tau)$ since g makes a negligible contribution. It is possible to demarcate the two regions satisfactorily because $P_h(\tau)$ itself consists (when ρ^2 is large) of a sequence of isolated and randomly occurring peaks. The first region will determine the accuracy with which range can be measured, whilst the area under the second region represents the degree of *ambiguity* of the observation. It is the probability that, in the ensemble of observations with fixed received waveform, the actual range does not lie anywhere near τ_0.

The more important features of the signal and noise functions have now been enunciated, and it has been seen that, if the total received energy is not large compared with the noise power per unit bandwidth, only a negligible quantity of information can be obtained from the system. This result is not, of course, confined to radar and may be seen quite generally from Shannon's formula, equation (1).

§ 5. RANGE ACCURACY.

There are two distinct statistical approaches to the problem of observational error. The first (and here deceptively simple) approach is to fix the true value and evaluate the spread of the estimates. The other, which has been adopted in this paper, is to fix on a typical received waveform resulting from a true range τ_0, and to consider the spread of the true values τ which could have produced it. The latter approach, involving the inverse distribution $P(\tau)$, eliminates consideration of how the observer estimates the true value when confronted with the received waveform.

Since $P(\tau)$ represents the properties of an ensemble in which the received waveform γ is fixed, it is *particular*, inasmuch as it depends on the specific choice of γ. It will therefore be necessary, on occasion, to consider average statistical properties of $P(\tau)$ by varying γ while keeping τ_0 fixed (see equation (9)). Averages over this fixed-τ_0 ensemble will be distinguished from averages in the fixed-γ ensemble by using the symbol Av instead of a bar. We shall then demonstrate that $P_g(\tau)$ is approximately Gaussian near τ_0 and calculate its standard deviation.

For the present, any ambiguity arising from the behaviour of $h(\tau)$ outside the g-region around τ_0 is ignored, but its influence inside it is interesting and is considered as follows. By equation (31) and the Taylor expansion of $u(t-\tau)$, the noise function near τ_0 is

$$h(\tau)=\frac{\alpha_0}{N}\{\mathbf{n}^\dagger\mathbf{u}(\tau_0)-(\tau-\tau_0)\mathbf{n}^\dagger\mathbf{u}'(\tau_0)+O(\tau-\tau_0)^2\}. \quad . \quad . \quad (35)$$

It may be shown that, if ρ^2 is sufficiently large, neglect of all but the linear and constant terms is not serious in spite of the necessity for including the quadratic term in the corresponding expansion of $g(\tau)$. Thus we may write

$$h(\tau)=\xi+i\eta=\{\xi_0+(\tau-\tau_0)\xi'\}+i\{\eta_0+(\tau-\tau_0)\eta'\}, \quad . \quad . \quad (36)$$

where ξ_0, η_0, ξ' and η' are random in the fixed-τ_0 ensemble and have Gaussian distributions about zero. From equation (32),

$$\mathrm{Av}(\xi_0^2)=\mathrm{Av}(\eta_0^2)=\rho^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

and by a similar argument, it is easy to show that

$$\mathrm{Av}(\xi'^2)=\mathrm{Av}(\eta'^2)=\rho^2\beta^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

Thus $|g+h|$ may be approximately evaluated in the neighbourhood of τ_0 for substitution into equation (24). Expanding the modulus and neglecting powers of $(\tau-\tau_0)$ higher than the second, one obtains from equations (30) and (36),

$$|g(\tau)+h(\tau)|=\rho^2+\xi_0+\frac{\eta_0^2}{2\rho^2}+\frac{\xi'^2}{2\rho^2\beta^2}-\tfrac{1}{2}\rho^2\beta^2\left(\tau-\tau_0+\frac{\xi'}{\rho^2\beta^2}\right)^2+O\left(\frac{1}{\rho^2}\right). \quad . \quad . \quad . \quad (39)$$

This may be written more simply as

$$|g+h|\approx\rho^2+\chi-\tfrac{1}{2}\rho^2\beta^2(\tau-\tau_m)^2, \quad . \quad . \quad . \quad . \quad (40)$$

where from equations (37) and (38)

$$\mathrm{Av}\,\chi=1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

$$\mathrm{Av}\,\tau_m=\tau_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

$$\mathrm{Av}(\tau_m-\tau_0)^2=1/\rho^2\beta^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Near τ_0, one therefore has from equation (24)

$$P_g(\tau)=\lambda I_0\{\rho^2+\chi-\tfrac{1}{2}\rho^2\beta^2(\tau-\tau_m)^2\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

The asymptotic expansion for I_0 is

$$I_0(x)\sim\frac{e^x}{\sqrt{(2\pi x)}}\left\{1+\frac{1}{8x}+O\left(\frac{1}{x^2}\right)\right\}, \quad . \quad . \quad . \quad . \quad (45)$$

and a trivial application of it shows $P_g(\tau)$ to be approximately Gaussian near τ_0 with

$$\tau=\tau_m, \qquad (46)$$

$$\overline{(\tau-\tau_m)^2}=1/\rho^2\beta^2. \qquad (47)$$

Thus the standard deviation of τ from its mean in $P_g(\tau)$ is given by

$$\delta\tau=1/\rho\beta, \qquad (48)$$

and this represents the *a posteriori* uncertainty in the measurement of range. That the mean value of τ in $P_g(\tau)$ has itself the same uncertainty with respect to τ_0 (equation (43)) is no mere coincidence, and is what one would expect from an application of inverse probability: both spreads really represent the same effect.

Equation (48) is the first result of practical interest. The uncertainty is independent of the frequency response of the receiver because it has been established without reference to any particular receiving mechanism and represents optimum performance.

§ 6. AMBIGUITY.

It was remarked at the end of § 4 that the degree of ambiguity of an observation is the probability that, in the ensemble of observations with fixed received waveform, the actual range τ does not lie anywhere near τ_0. The ambiguity, A, may therefore be expressed as the area under that portion of $P(\tau)$ which is due only to $h(\tau)$, *i. e.* which excludes the signal peak discussed in the previous section. This will be calculated by multiplying the expected area per unit range of τ due to h alone, by the whole of the *a priori* interval T. This in fact includes a small range of integration near τ_0 which should strictly be omitted, but since for any practical system one has the inequality

$$T \gg 1/\rho\beta \qquad (49)$$

this approximation is justified.

It is convenient to define $P_h(\tau)$ precisely, by the relation

$$P_h(\tau)=\lambda I_0(|h|), \qquad (50)$$

and this is identical with $P(\tau)$ defined in equation (24), except near τ_0. The "expected area per unit range" under $P_h(\tau)$ is, of course, the expected value of P_h and may be obtained by considering the probability distribution for $|h|$ given by equation (33). Thus

$$\mathrm{Av}P_h=\overline{P}_h=\int_0^\infty P_h Q(P_h)\,dP_h, \qquad (51)$$

where

$$Q(P_h)\,dP_h=q(|h|)\,d|h|. \qquad (52)$$

The two forms of average in (51) are equal because $h(\tau)$ is a stationary random function. The integral may be evaluated by substituting for q from equation (33), P_h from equation (50), and using the expansion (45) for the Bessel function. Finally, one obtains

$$\overline{P}_h=\lambda e^{\rho^2/2}(1+O(\rho^{-4})), \quad \ldots\ldots\ldots \quad (53)$$

and the ambiguity, given by the average area under $P_h(\tau)$ is

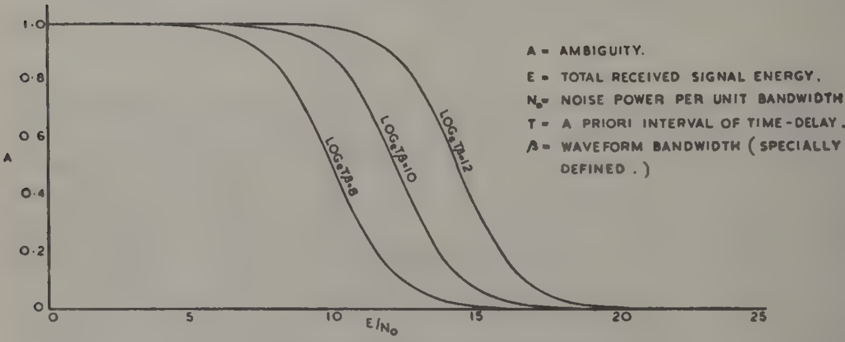
$$A=T\overline{P}_h. \quad \ldots\ldots\ldots \quad (54)$$

It only remains to determine the normalizing constant, λ, of $P(\tau)$, which depends also on the area under $P_g(\tau)$ due to the signal function.

This area is obtained by integrating the distribution given by equation (44), and using again the Bessel expansion. The result is*

$$\int_T P_g(\tau)d\tau=\frac{\lambda e^{\rho^2+1}}{\rho^2\beta}\left\{1-\frac{1}{8\rho^2}+O\left(\frac{1}{\rho^4}\right)\right\}. \quad \ldots \quad (55)$$

Fig. 1.



The threshold of intelligibility.

Neglecting terms of order ρ^{-4} compared with unity, the normalizing equation is

$$1=\int_T P_g(\tau)\,d\tau+\int_T P_h(\tau)\,d\tau$$
$$=\lambda\left\{\frac{e^{\rho^2+1}}{\rho^2\beta}\left(1-\frac{1}{8\rho^2}\right)+Te^{\rho^2/2}\right\}. \quad \ldots\ldots \quad (56)$$

Hence, approximately,

$$A=\frac{T\rho^2\beta}{T\rho^2\beta+\exp\left(\frac{1}{2}\rho^2+1\right)}. \quad \ldots\ldots\ldots \quad (57)$$

---

* An average has been adopted in the exponent—see equation (41)—as a matter of expedience. The consequences in later parts of the theory are unimportant except when $A\approx 0{\cdot}5$, where additional approximations are, in any case, unavoidable.

Thus, keeping T and β constant, as the ratio $\frac{1}{2}\rho^2$ of the received signal energy to the noise power per unit bandwidth increases, the ambiguity diminishes to zero from a value near unity. The transition is fairly abrupt as shown in fig. 1, and forms the threshold of intelligibility. This may be defined conventionally by $A=0{\cdot}5$, and is therefore given by

$$\tfrac{1}{2}\rho^2+1=\log T\rho^2\beta. \quad \text{. . . . .} \quad (58)$$

When $\frac{1}{2}\rho^2+1 \gg \log T\rho^2\beta$, unambiguous reception is ensured. This is the second criterion for a useful radar observation.

§7. INFORMATION.

The quantity of information obtainable from a radar observation is the difference of the entropies (Shannon 1948) of the *a priori* and *a posteriori* probability distributions for range. Since the *a priori* distribution has been assumed rectangular of width T, its entropy, defined by *

$$H=-\int p \log p \, d\tau, \quad \text{. . . . .} \quad (59)$$

is given by

$$H_0=\log T. \quad \text{. . . . .} \quad (60)$$

The *a posteriori* entropy required is the average entropy of $P(\tau)$ over the total ensemble of received waveforms. Since, however, the entropy is independent of τ_0, it is sufficient to consider an average over the fixed-τ_0 ensemble. As before, that portion of the distribution near τ_0, viz. $P_g(\tau)$, and the remainder, $P_h(\tau)$, will be considered separately, the total entropy being given by addition.

A difficulty arises, however, in evaluating the entropy of any particular distribution $P(\tau)$ because, in doing so, the correct normalizing constant for that distribution should be used, and this will vary throughout the fixed-τ_0 ensemble. This makes the integrations difficult to perform, and to simplify the treatment a fixed normalizing constant λ, dependent on average properties of $P(\tau)$ and given by equation (56), is used.

Fluctuations in λ are due, when $\rho^2 \gg 1$, almost entirely to $P_h(\tau)$ and their neglect is justified in either of two circumstances:

(*a*) When the total area under $P_h(\tau)$ is small, *i. e.* $A\sim 0$, and the entropy is approximately that of $P_g(\tau)$ alone, or

(*b*) When the total interval, T, over which $P_h(\tau)$ extends, is sufficiently large for P_h to include an adequate statistical sample of those of its ensemble properties which influence the entropy. This condition amounts to $A\sim 1$, though the derivation is lengthy and is here omitted. The entropy then arises almost entirely from $P_h(\tau)$.

Thus one may first calculate the average entropy due to $P_g(\tau)$ alone, which will represent that of $P(\tau)$ when $A\sim 0$; then that due to $P_h(\tau)$ alone, giving the entropy of $P(\tau)$ when $A\sim 1$; finally, an approximation in the intervening (threshold) region will be made by ignoring fluctuations of the normalizing constant.

---

* Here $p(\tau)$ signifies any probability distribution.

The contribution of $P_g(\tau)$ to the entropy, is obtained by substituting the averaged expression for $P_g(\tau)$ given by equations (44) and (41) into equation (59), using the expansion for I_0, and integrating. This leads to

$$Av(H_g)=-\frac{\lambda e^{\varrho^2+1}}{\rho^2\beta}\left\{\left(1-\frac{1}{8\rho^2}\right)\log\frac{\lambda}{\rho\sqrt{(2\pi)}}+\rho^2+\tfrac{3}{8}+O\left(\frac{1}{\rho^2}\right)\right\}. \quad . \quad (61)$$

But if $A\sim 0$, from equation (55) the normalizing constant is given by

$$\frac{\lambda e^{\varrho^2+1}}{\rho^2\beta}\left(1-\frac{1}{8\rho^2}\right)=1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

whence from equation (61)

$$AvH_g=\log\frac{\sqrt{(2\pi e)}}{\rho\beta}+O\left(\frac{1}{\rho^2}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

The information content is given by the difference of the *a priori* and *a posteriori* entropies, viz.

$$I=H_0-H_1=\log T-H_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (64)$$

and hence

$$I\sim\log\frac{T\rho\beta}{\sqrt{(2\pi e)}}, \quad (A\sim 0). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

Since $P_h(\tau)$ is a stationary random function, one may calculate its average contribution to the total entropy by multiplying the expected value of $P_h \log P_h$ by the interval T of integration. Thus, ignoring fluctuations of λ,

$$AvH_h=-T\int_0^\infty Q(P_h)P_h\log P_h dP_h. \quad . \quad . \quad . \quad . \quad (66)$$

Hence, substituting for $Q(P_h)$ and P_h from equations (52), (33) and (50), using the expansion for I_0, and integrating, one obtains

$$AvH_h=-\lambda Te^{\varrho^2/2}\left\{\log\frac{\lambda}{\rho\sqrt{(2\pi)}}+\rho^2+\tfrac{1}{2}+O\left(\frac{1}{\rho^2}\right)\right\}. \quad . \quad . \quad (67)$$

But when $A\sim 1$, using equations (53) and (54), the normalizing constant λ is given approximately by

$$\lambda e^{\varrho^2/2}=1/T. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

Thus,

$$AvH_h=\log(T\rho\sqrt{(2\pi)})-\tfrac{1}{2}\rho^2-\tfrac{1}{2}+O\left(\frac{1}{\rho^2}\right) \quad . \quad . \quad . \quad (69)$$

and

$$I\sim\tfrac{1}{2}\rho^2-\log(\rho\sqrt{(2\pi)})+\tfrac{1}{2}, \quad (A\sim 1). \quad . \quad . \quad . \quad . \quad (70)$$

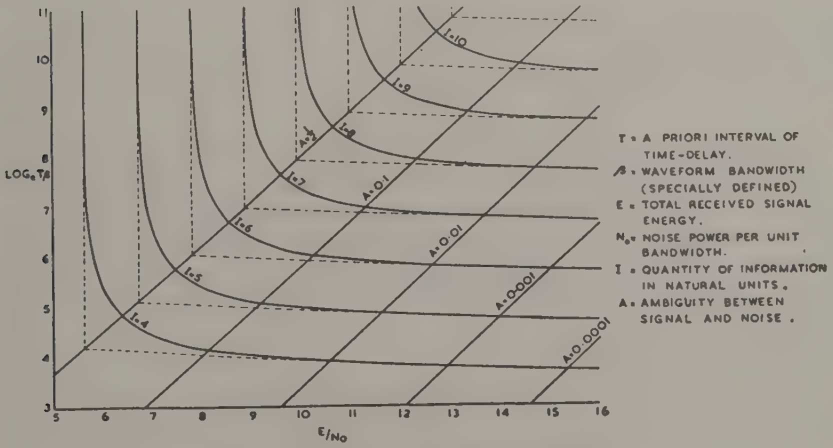
Equations (65) and (70) thus give the asymptotic approximations to the information content under conditions of unambiguous and ambiguous reception, except in the vicinity of the threshold at $A=0{\cdot}5$.

When A approximates neither to zero nor unity, it is, of course, necessary to normalize $P(\tau)$ taking into account both $P_g(\tau)$ and $P_h(\tau)$, and incidentally to ignore—as explained above—normalizing fluctuations which should in this region be included. The value of λ taken from equation (56) is substituted into equations (61) and (67), addition of which gives the approximate average *a posteriori* entropy H_1. Finally subtraction from H_0 gives

$$I \approx -\log\left\{\sqrt{(2\pi e)}\left(\rho e^{-\frac{1}{2}\rho^2-1}+\frac{1}{T\rho\beta}\right)\right\}+O\left(\frac{1}{\rho^2}\right). \quad . \quad . \quad (71)$$

It will be seen that this expression reduces, as it must, to equations (65) or (70) according as the second or first terms in the logarithm predominate, *i.e.* A∼0 or 1. It represents a better approximation to the total quantity of information than either of the previous equations, even though fluctuations of the normalizing constant are ignored.

Fig. 2.



Information and ambiguity contours.

§8. Discussion.

The chief results of the foregoing theory are illustrated in fig. 2, where contours of information content I, given by equation (71), and ambiguity A, given by equation (57), are plotted against the quantities $\log_e(T\beta)$ and E/N_0. It will be seen that the I-contours behave in a markedly different manner on either side of the threshold of intelligibility (A=0·5). When the ambiguity A is small, the I-contours are asymptotic to dotted curves given by equation (65). The information here represents almost unambiguous knowledge of target range, subject to a "topological" error which is shown in §5 to have a standard deviation of

$$\delta\tau = 1/\rho\beta,$$

where

$$\rho = \sqrt{(2E/N_0)}$$

and β is the "effective bandwidth" of the transmitted waveform. As would be expected, the error diminishes as the total received energy E increases, if the noise power N_0 per unit bandwidth is constant. It also diminishes as β increases, but increasing β alone in order to obtain high resolution, has its drawbacks because the observation gradually becomes more ambiguous.

When the ambiguity is greater than a half, the I-contours are asymptotic to the dotted vertical lines given by equation (70), which represent the upper bound to the information content when the total received energy is fixed. In Shannon's general theory of communication (Shannon 1948) there is a similar upper bound, with which comparison may be made. Quite generally, the quantity of information I which can be extracted from any received message of energy E is limited to E/N_0 natural units. This result is obtained from equation (1) by writing

$$I=Ct=Wt\log\left(1+\frac{E}{N_0Wt}\right)\text{nat. units},$$

where t is the duration of the message. As $Wt\to\infty$, I approaches the asymptotic value

$$I_s=E/N_0=\tfrac{1}{2}\rho^2.$$

The corresponding radar limit, as $\beta\to\infty$, is less than this by $\log(\rho\sqrt{(2\pi)})-\frac{1}{2}$ natural units of information, and the discrepancy may be shown to correspond to the fine-structure information which was deliberately discarded in passing from equation (23) to (24). If radar could operate without a high-frequency carrier, this effect would not arise and the asymptotes in the ambiguous region of fig. 2 would exactly equal Shannon's theoretical limit.

In normal reception, the quantities β, N_0, and the *a priori* interval T are constant, while E increases with the time of observation. One therefore starts beyond the left of the diagram and moves steadily to the right along a horizontal path. At first, the amount of information increases rapidly as the ambiguity is steadily removed. After crossing the threshold into the useful region, however, the information increases only very slowly and represents a progressive improvement in accuracy. This phenomenon is in full accord with the ideas underlying information theory. Once the approximate position of the target is known without ambiguity, further incoming energy is continually contributing information which is already partly known; the resolution in range is increased by continued emphasis rather than by a systematic process of elimination. For this reason, the most efficient radar system should work as close to the threshold of intelligibility as possible—how close depending upon the degree of ambiguity which can be tolerated.

It must be emphasized that a phase-coherent received waveform is assumed in this paper, and this severely limits the practical application of the theory. If the range parameter τ were itself a function of time, the time-integrals in the theory would have to be evaluated differently. As matters stand, the total period of observation must be restricted to an interval which is sufficiently short for τ to be treated as a constant. Under these circumstances, there is nothing to choose (except on practical grounds) between concentrating the available energy into one repetition period of the modulation or spreading it out over several. There is little doubt, however, that the latter course is wasteful of energy when long-term phase-coherent integration cannot be utilized (Lawson and Uhlenbeck 1950).

The bandwidth of the receiver does not appear in the results of the theory. As long as it is sufficiently wide to include the whole of the transmitted spectrum, no information is gained or lost by increasing it, and none can be gained by reducing it. The choice of frequency response at the receiver depends only on the best way to extract, and present to the observer, all the latent information in the incoming waveform, with which practical problem this paper is not concerned.

§9. ACKNOWLEDGMENT.

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this paper.

§10. REFERENCES.

GABOR, D., 1946, *Jour. Inst. Elect. Eng.*, **93**, Pt. III, 429.
JEFFREYS, H., 1939, *Theory of Probability* (Oxford).
LAWSON, J. L., and UHLENBECK, G. E., 1950, *Threshold Signals*, Chap. 7.
OSWALD, J., 1949, *Comptes Rendus* (l'Académie des Sciences), **229**, 21.
RICE, S. O., 1945, *Bell. Sys. Tech. Jour.*, **24**, 46.
SHANNON, C. E., 1948, *Bell Sys. Tech. Jour.*, **27**, 379 and 623 ; 1949, *Proc. Inst. Radio Eng.*, **37**, 10.
TITCHMARSH, E. C., 1939, *Theory of Fourier Integrals* (Oxford).
WIENER, N., 1949, *Interpolation, Extrapolation, and Smoothing of Stationary Time Series.*

XCIII. *Some Optical Properties of Evaporated Layers of Silver, Copper and Tin.*

By D. G. Avery, Ph.D.,
Royal Holloway College, London University*.

[Received July 21, 1950.]

[Plates XXXII. & XXXIII.]

Summary.

In an earlier work (Tolansky 1944), a description was given of the use of multiple-beam interferometry to determine relative phase changes on reflection at the silvered surfaces of an interferometer. This method has been extended and critically examined. It is shown that the method does not approach the accuracy of the older polarimetric methods, and that its optimum use is restricted to reflecting surfaces of fairly high reflection coefficient. Within its limitations the method has advantages.

Measurements of the relative phase changes on reflection at a number of layers of silver, copper, and tin have been made, together with measurements of the ellipticity of the light reflected at these layers. The variation of these quantities with the thickness of some layers of silver has been discussed in relation to the use of evaporated layers for the measurement of the optical constants of bulk materials, and the optical constants of evaporated layers of copper and tin have been determined and the results discussed.

Part I: The Interferometric Technique for the Measurement of Relative Phase Changes.

In 1944 Tolansky described the appearance of the multiple-beam interference fringes formed by a highly silvered lens and plate combination, when the angle of incidence of monochromatic light falling on to the interferometer varied from 0° to 80°. At normal incidence the fringes took the form of rings (Newton's rings), but as the angle of incidence increased these of course became elliptical, but further each became a doublet, the separation of the doublet components increasing with the angle of incidence. The components were found to be polarized in mutually perpendicular directions and the origin of the doubling was shown to be the effective difference in gap of the interferometer for the two components, due to the difference between the phase changes suffered on reflection by the light polarized parallel and perpendicularly to the plane of the incidence.

---

* Communicated by Professor S. Tolansky.

It was apparent that this was a new method for the measurement of the relative phase changes occurring on reflection, and Tolansky measured this quantity for the particular silvered surfaces of his interferometer. No further examination of this method has been reported and it is the purpose of this paper to study the method critically both theoretically and in the light of a number of measurements of relative phase changes made by this technique.

The Theory of the Method.

Although Tolansky's original experiments were carried out on silver-modified Newton's rings it is advantageous to treat the simpler case of straight line fringes formed in a wedge interferometer composed of two plane reflecting surfaces, inclined at a small angle.

Consider first monochromatic light, of wavelength λ, incident at angle θ on a wedge interferometer with an air gap. The fringes (of equal thickness) will take the form of lines parallel to the line of intersection of the two reflecting surfaces. If the angle of incidence is not zero, each fringe will in fact have two components, and if the fringes occur at distances x_s, x_p x'_s, x'_p, etc. along the wedge, it can be shown that the relative phase change Δ can be determined from

$$\Delta(\theta, \lambda)=\lambda/2 \,.\, (x_s-x_p)/(x'_p-x_p). \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

An equation from which Δ may be determined can also be derived in the case of fringes of equal chromatic order (Tolansky 1945). These fringes are formed when an interferometer is illuminated with white light, and a small section of the interferometer is projected on to the slit of a spectrograph. The section is small enough for the value of t to be assumed constant over it, if the wedge apex is parallel to the spectrograph slit. In the focal plane of the spectrograph fringes appear, which in this case take the form of straight lines parallel to the slit. Again, if θ is not zero, the fringes are doubled, the components having different wavelengths.

Let the wavelengths of the components of two adjacent doublets be λ_p, λ_s, λ'_p, λ'_s. Then the relative phase changes given by

$$\Delta(\theta, \lambda)=m/2(\lambda_p-\lambda_s), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and hence Δ can be found from measurements of λ_p and λ_s if m, the order of the fringe, is known.

The order, m, is given by

$$m\lambda_p-(m+1)\lambda'_p=2\Phi_p(\theta, \lambda_p)-2\Phi_p(\theta, \lambda'_p) \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where the function Φ_p represents the change of phase suffered on reflection by the parallel component. In the absence of published data describing the variation of the function with wavelength and with angle of incidence, the order of the fringe m cannot be determined without making an assumption of the form of Φ_p. This is true not only for fringes at other than normal incidence but also for all uses of fringes of equal chromatic order.

If we assume $2\Phi_p(\theta, \lambda)=\beta\lambda$ where β is a constant fraction, then as, closely enough,

$$\lambda_p \simeq \tfrac{1}{2}(\lambda_p+\lambda_s) \equiv \lambda, \ldots$$

gives

$$m-\beta=\lambda'/(\lambda-\lambda'). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

Now m, the order of the fringe, is integral, and the quantity $m-\beta$, which can be determined from measurements of the mean wavelengths λ and λ', is in general not integral. Moreover, the sign of the fractional part is indeterminate.

However, from measurements on fringes of equal thickness we can find the relative phase change Δ for one wavelength unambiguously from equation (1), and so if the value of m for a fringe of equal chromatic order at that wavelength be chosen to give the correct value for Δ, the values for the orders of the other fringes of equal chromatic order at different wavelengths can readily be assigned, and the fringes used to measure the variation of the relative phase change with wavelength. In addition fringes of equal chromatic order, being formed in a very small section, of the interferometer of assumed constant thickness, have a constant dispersion leading to improved accuracy of measurement. As against this latter point however, it should be noted that when fringes of equal thickness are used, the whole surface of the interferometer is covered, with perhaps a desirable averaging effect of the properties of the films.

The experimental arrangement and the interferometric technique have been adequately described elsewhere (Tolansky 1948) and no treatment will be given here. The discussion which follows of the accuracy of measurement of relative phase changes from measurements of the doubling of interference fringes of non-normal incidence will assume that the interferometer has been used in such a manner as to give the best interference fringes obtainable with the particular reflecting surfaces employed.

The Errors in Measurement.

In figs. 1 (*a*), (*b*) and (*c*) (Pls. XXXII. & XXXIII.) are shown photographs of fringes formed at varying angles of incidence in interferometers whose surfaces had reflection coefficients 90, 50, and 20 per cent respectively. The photographs are all of fringes of equal thickness, using light of wavelength 5461 Å.

The previously established increase in separation of the fringe doublet as the angle of incidence increases from zero is clearly shown.

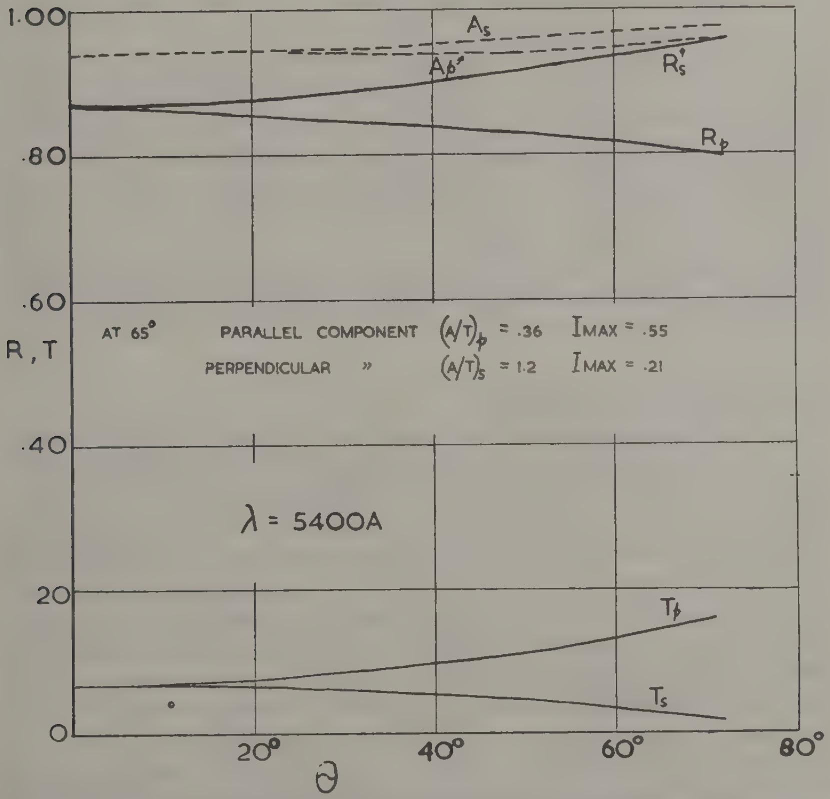
Three striking features of these fringes are immediately apparent, all of which are of importance in assessing the accuracy with which the doubling can be measured, and the relative phase change determined.

First, comparison of figs. 1 (*a*), (*b*) and (*c*) (Pls. XXXII. & XXXIII.) shows that the width of both components of the fringe doublets increases rapidly with decreasing reflection coefficient. This of course is to be expected and constitutes one serious drawback to the generalized use of the technique, since the highest accuracy is only obtainable when measuring the effect on surfaces of reflection coefficients of over 85 per cent.

Secondly, as the angle of incidence increases there is a considerable difference in fringe width for the two components for all values of the reflection coefficient, and thirdly, the maximum intensity of the sharper component decreases rapidly at the higher angle of incidence.

Tolansky suggested that the difference in the intensities was evidence for the existence of some preferential absorption by the silver film. However, reference to fig. 2, which shows the variation with incidence angle of the reflection and transmission coefficients of a typical silver film, shows that for these films, whilst some slight preferential absorption exists at the higher angles, it is in the opposite sense to the effect observed. The difference in the maximum intensities of the two components arises mainly from the difference in the transmission coefficients in the two cases, the maximum intensity being governed by the ratio $1/1+(A/T)^2$ where A is the absorption given by $A=1-(R+T)$. The values shown in fig. 2 are typical for silver films of high reflection coefficient. The

Fig. 2.



The variation with incidence angle of the reflection and transmission coefficients of a typical silver film.

differences in fringe width are of course due to the widely differing values of the reflection coefficient for the components. This effect increases with decreasing normal incidence reflection coefficient.

The accuracy with which the doublet separation can be measured is affected by the differences in width and intensity of the components in two ways. Not only is it impaired by the increase in fringe width due to the lower reflection coefficient of the "parallel" component, but also this width, coupled with the weaker intensity of the "perpendicular" component, causes the visibility of the latter to be poor. This obscuring of the weaker component is not serious for high reflection coefficient films, but becomes so when the reflection decreases to values of 50 per cent or less (fig. 1 (*c*), Pl. XXXIII.).

The errors all increase with decreasing reflection coefficients. With fringes of equal thickness and fringes of equal chromatic order, in transmission, determinations of relative phase change have been made for surfaces of reflection coefficient varying between 20 and 94 per cent. The experimental error in the determination of relative change of phase, expressed as a fraction of the wavelength of the incident light, varies from $\pm 0{\cdot}002\,\lambda$ for the optimum case of $R > 80$ per cent, to $\pm 0{\cdot}005\,\lambda$ for $R < 30$ per cent. If the relative phase change is expressed as an angle this corresponds to $\pm 0{\cdot}7°$ for the optimum, and $\pm 2°$ for the worst case studied.

The Limitations of the Technique.

The accuracy obtainable with this method for measuring relative phase changes ($\pm 0{\cdot}7°$) is inferior to the figures of $\pm 0{\cdot}001\,\lambda = 0{\cdot}36°$ readily obtainable by Drude's method (Hass, 1938) and with the error of $\pm 0{\cdot}1°$ in the estimation of the angle of principal incidence by O'Bryan (1936).

From the point of view of the use of the measurements for the determination of optical constants, however, the higher accuracies obtainable in the older optical methods are not always required, since large local irregularities often arise in the preparation of the reflecting surfaces employed. Clearly under such conditions high intrinsic accuracy of measurement is meaningless physically. It was to overcome these errors that the use of evaporated layers as characteristic reflecting surfaces of metals was introduced (O'Bryan, Hass, *loc. cit.*). However, this technique does not allow of one main advantage of the use of evaporated layers to be gained, namely the measurement of the properties of such layers *in vacuo*. Possible deterioration of the layers by atmospheric corrosion during measurement could be avoided by the introduction of an oil between the interferometer plates.

However, a principal limitation of the method remains. Relative phase changes occurring on reflection at low reflectivity metals such as copper, tin, chromium, etc. can only be measured with half the accuracy available for measurements on silver and aluminium, due to the increase of the errors with decreasing reflection coefficient.

As against these limitations must be set some advantages over the earlier method. The method requires no expensive polarimetric equipment and the experimental arrangement is simple. The doubling of the fringes can be photographically recorded and measurements made at leisure, as opposed to the need for laborious visual observations necessary with most of the polarimetric methods. This is particularly the case when it is desired to measure the variation of the relative change of phase with wavelength, since the whole visible range can be covered with a single photograph of fringes of equal chromatic order, whereas a dozen or more visual observations with a half-shade device would be necessary with the older methods.

PART II : THE OPTICAL MEASUREMENTS.

The interferometric data are coupled with measurements of the ellipticity of the light reflected at the layers, made with a simple photoelectric photometer. An investigation was made as to whether the magnitude of the effects occurring on reflection at evaporated layers having the finite transmission required for the multple-beam method were the same as those associated with reflection at the bulk metal. The variation of the phase and intensity quantities with the thickness of the evaporated layers has been measured to supplement existing information regarding the normal incidence reflection characteristics of evaporated layers.

Measurements have been made for a number of evaporated layers of silver of varying thickness and the results served to define the limits within which the thickness of a metallic layer must lie if it is to be used for the determination of the optical constants of the metal. With the increasing use of evaporated layers in this connection (*e. g.* O'Bryan 1936, Hass 1938, Archard 1949) these limits are a matter of some significance, particularly when the interference method is used for the measurement of the relative phase changes.

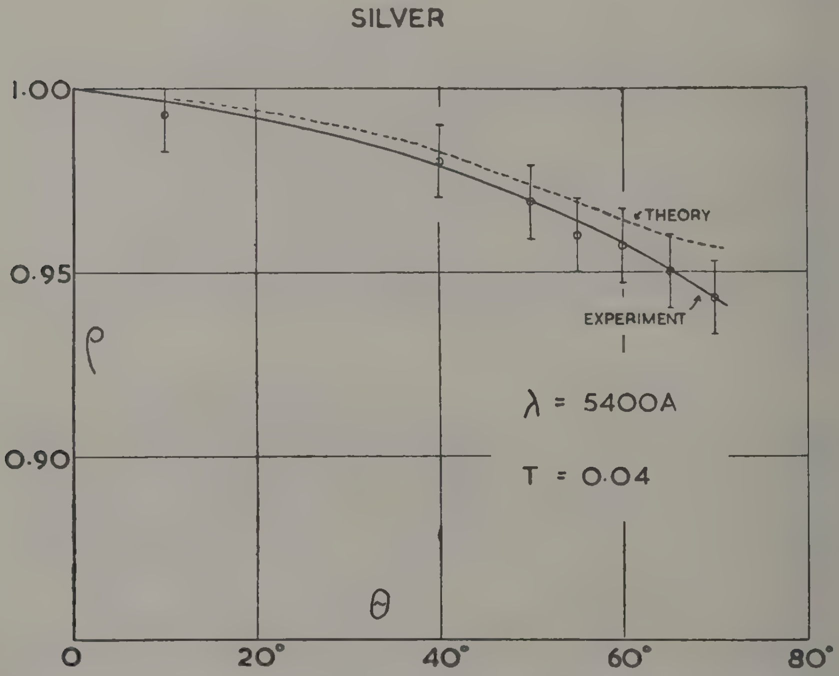
The layers of silver, and of the other metals used, were all produced by vacuum evaporation from molybdenum filaments at pressures less than 5×10^{-5} mm. of Hg, after the glass disks on to which they were deposited had been cleaned both by hand and by ionic bombardment in the evaporation chamber. The rates of evaporation were such that opaque layers of the metals could be formed in thirty seconds.

Measurement of Relative Phase Changes and Ellipticities for Reflection at Silver Layers.

The first series of silver layers produced were all of the order of thickness used for usual interferometric work, *i. e.* between 400–500 Å. thick and having transmission coefficients of up to 4 or 5 per cent. The variation of the ellipticity ρ with the angle of incidence of light for one of these layers is shown in fig. 3, and the variation of the relative phase changes with incidence angle for several layers in fig. 4. The error in the measurement of the ellipticity throughout the work was $\pm 0{\cdot}01$, the error in the measurement of the relative phase change, Δ, varied with the reflection coefficient of the layers, and for the present series was $\pm 0{\cdot}002\,\lambda$. The measurements of ellipticity were made for the light passed by a filter giving a mean wavelength of 5400 Å. and a band width of roughly 300 Å., and of relative phase change for the mercury green line 5461 Å.

Taking the values found by Hass (1938) for the optical constants of silver ($n=0{\cdot}177$, $k=18{\cdot}60$, $\lambda=5461$ Å.) the ellipticities and relative phase changes to be expected on reflection at a silver surface have been calculated using formulæ due to Drude (1890). These theoretical values have been plotted in figs. 3 and 4, and in addition, the angle of principal incidence for silver (for which the relative phase change on reflection is equal to $\lambda/2$ or $0{\cdot}25$) found by O'Bryan (1936) is also shown in fig. 4.

Fig. 3.

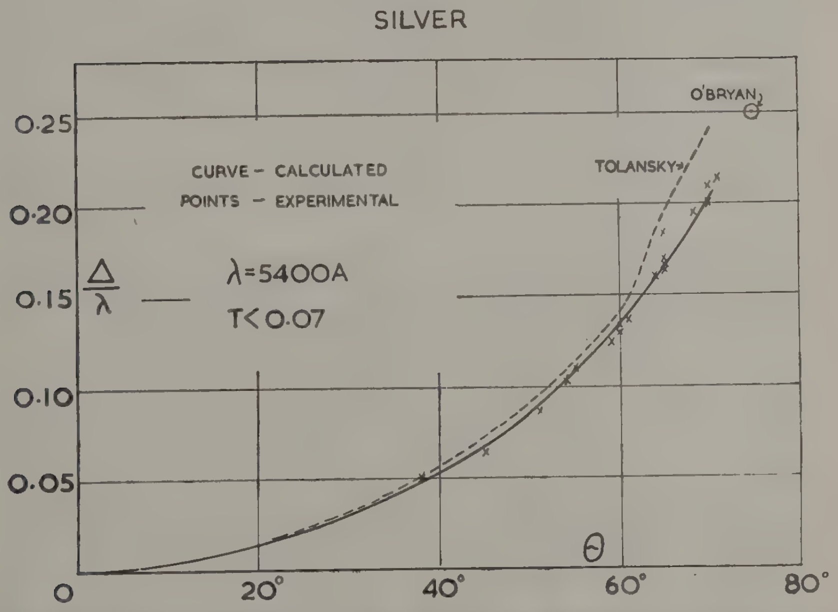


The variation with incidence angle of the ellipticity of the light reflected from a silver film.

It is seen from these curves that whilst the ellipticity for a layer having a transmission coefficient of 0·04 differs from the theoretical value for reflection at bulk silver by nearly one per cent at the higher angles of incidence, the relative phase change values, taken for some ten layers having transmission coefficients between 0·01 and 0·07 do not differ from the theoretical value to this extent. This indicates that the value of the ellipticity of the light reflected from evaporated layers is more sensitive to film thickness than is the relative change of phase on reflection. This will be borne out completely when the measurement of the variation of these quantities with thickness is considered.

In fig. 4 are plotted the earlier measurements made by Tolansky (1944) for the relative phase changes occurring on reflection with a given silver layer. In the neighbourhood of 65° incidence, there is a sudden change in the slope of Tolansky's curve. This effect has not been observed with any of the films examined in this work, and has not been completely explained. One possible explanation is that the films used by Tolansky were comparatively old (at least one month after preparation) and the effect may be due to surface contamination of the layer due to age. In the present work measurements were made on the layers within one hour of production in each case. Below 65° there is reasonably close agreement between the two sets of data.

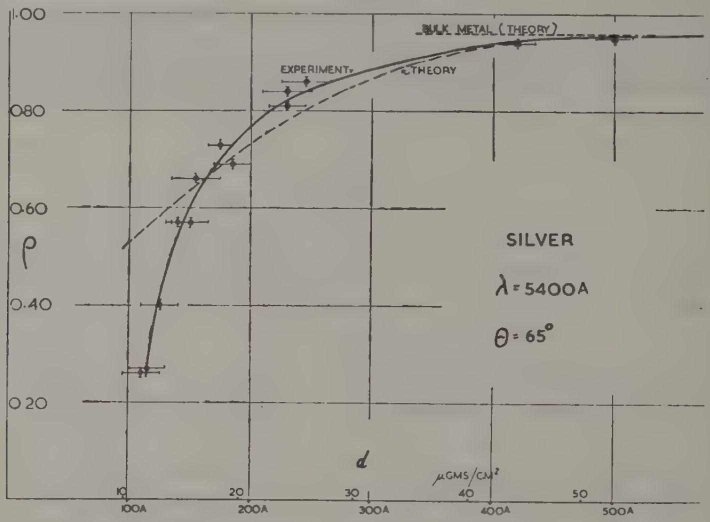
Fig. 4.



The variation of the relative phase change on reflection at "thick" silver films with the angle of incidence.

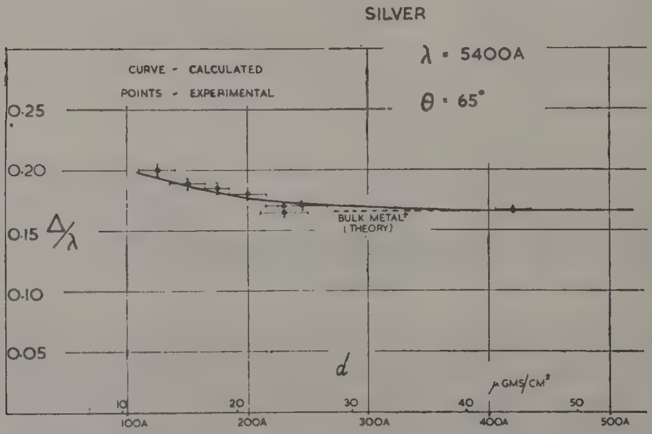
The variations of the quantities ρ and Δ with the thickness of the silver layer varying between 100 Å. and 500 Å. are shown in figs. 5 and 6 for an angle of incidence of 65°. The errors in the values for ρ are $\pm 0{\cdot}01$, but the errors in the relative phase changes increase from $\pm 0{\cdot}002\,\lambda$ for the thickness layers, to $\pm 0{\cdot}005\,\lambda$ for the thinnest ones, due to the decreasing reflection coefficient of the layers. The thicknesses of the layers were measured by the interferometric technique developed by Khamsavi and Donaldson (1947), and the values are accurate to ±15 Å. The mass of silver present per unit area of the film is also given, since the thickness measurements may be open to question (Avery 1949).

Fig. 5.



The variation of ellipticity with thickness for silver films.

Fig. 6.



The variation of relative phase change with thickness for silver films.

The theoretical values of ρ and Δ were calculated using formulæ similar to those developed by Winterbottom (1948).

Apart from a slight divergence in the thickness region about 250 Å., the values of the ellipticity measured agree with the theoretical values down to layers of about 150 Å. in thickness, and marked differences between theory and experiment first appear at thicknesses below this value. This agrees well with the variation of the normal incidence coefficients with thickness as measured by a number of workers (Rouard 1949), and the deviation of experiment from theory at the lower thickness values is in accord with the observed changes in the optical constants of silver layers of lesser thickness. The relative phase changes (fig. 6) show no similar divergence, though the measurement of d for the thinner films is correct only to 5 per cent.

If reflection at an evaporated layer is treated as equivalent to reflection at the metal in bulk when the values of ρ and Δ do not differ from the values characteristic of the bulk metal by more than one per cent, then fig. 6 shows that as far as the measurement of relative phase change is concerned, the layers need be no more than 350 Å. thick for this. For silver this corresponds to a transmission coefficient of 6 to 8 per cent. Such a thickness gives bright good multiple-beam interference fringes. Since the deviation of the value of Δ from the bulk metal value as the thickness decreases is probably due to interference effects within the layer, this implies that the optical transmission of the layer is the deciding factor of the usefulness of the film for " bulk " measurements, not the physical thickness.

Fig. 5 shows that the variation of the ellipticity with thickness is more rapid, and deviations of one per cent from the bulk metal value are found for thicknesses below 450 Å., that is for silver, for transmissions of no more than one or two per cent. Whilst for silver this represents no hindrance to the use of the same layers for the measurement of both the quantities ρ and Δ, metals which form layers of a greater absorption than silver layers need a greater transmission than this if reasonably intense interference fringes are to be obtained.

If a simple photometer to measure ρ, combined with an interference technique to measure Δ, is used to determine optical constants, or for other reflection studies, it seems that it is most suitable to prepare two sets of layers. One set should be opaque or very nearly so, and used for the measurement of the ellipticity, and the other set should have transmissions of the order of five per cent, and used in the interferometer to measure the relative phase changes. Where a polarimetric method is being used to determine ρ and Δ in one operation and the layers are produced and measured immediately *in vacuo*, the layers must have transmission of less than one or two per cent for the layer characteristics to be within one per cent of those of the bulk metal.

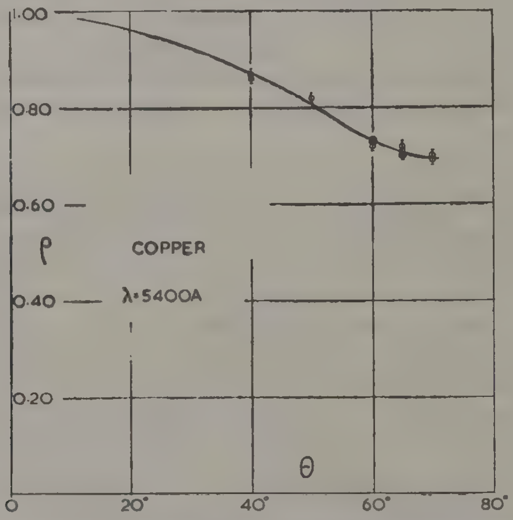
Electron microscope investigations have shown (Levinstein 1949) that not all metals can be deposited on to substrates at room temperature to give layers with structures corresponding to the bulk metal. Thus the optical properties of evaporated layers, particularly those of the low melting point metals, may differ from those of the bulk metal due to differences in crystalline structure, even in layers thick enough for the interference effects within the films to be neglected.

The Optical Constants of some Evaporated Layers of Copper.

Copper films were deposited by evaporation, four layers being prepared almost opaque and four with transmissions of up to five per cent. The values of the ellipticity found from the thick layers are plotted against incidence angles in fig. 7 and the values of relative phase change found interferometrically from the slightly transparent layers are shown in fig. 8. The errors in Δ are $\pm 0{\cdot}003\,\lambda$ due to the relatively low reflection coefficient of copper (0·57 at normal incidence).

Taking the values of ρ and Δ for four angles of incidence, 50°, 55°, 60°, 65°, the optical constants for the layers were determined using the formulæ of Drude. The value derived for the refractive index is $n=1{\cdot}10\pm0{\cdot}07$, and for the absorption coefficient $k=2{\cdot}3\pm0{\cdot}1$.

Fig. 7.



The variation of ellipticity with incidence angle for thick copper films.

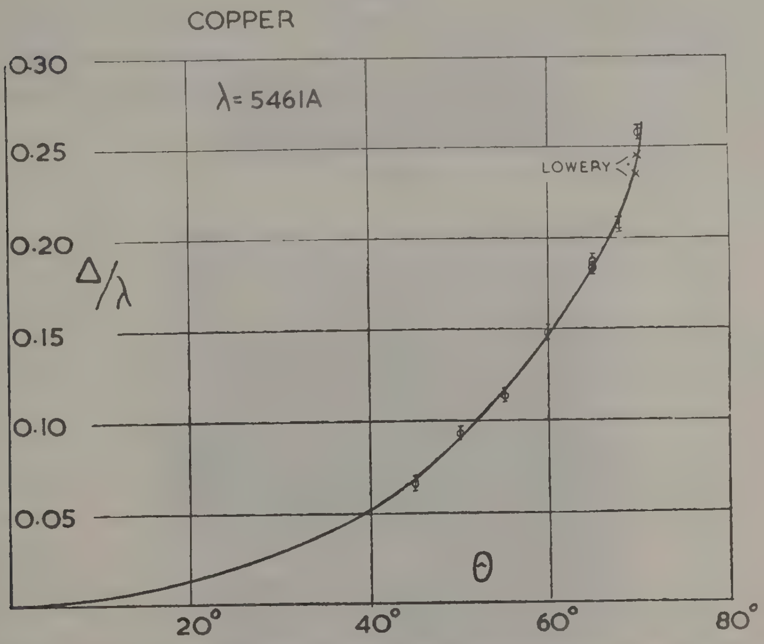
Also included in fig. 8 are two values for Δ for evaporated layes of copper reported by Lowery, Wilkinson and Sware (1936). Good agreement exists within the present values. However the values of the optical constants calculated by Lowery, $n=0{\cdot}90$, $k=2{\cdot}1$, for a single angle of incidence differ from those quoted above.

These are in close agreement with the values quoted in the *International Critical Tables* for $\lambda=5400$ Å., $n=1{\cdot}07$, $k=2{\cdot}25$, and with those given by Winterbottom (1936) for bulk copper, $n=1{\cdot}00$, $k=2{\cdot}40$. However, as Lowery has demonstrated, values found from measurements on copper surfaces prepared by polishing are extremely susceptible to differing methods of surface preparation, hence the agreement is fortuitous. The values for electrolytically deposited copper are $n=1{\cdot}38$, $k=1{\cdot}78$, and for cathodically sputtered layers $n=0\cdot90$, $k=2\cdot70$, and so it seems that the layers produced by evaporation resemble the structure of the massive metal more closely than layers made by electrolytic deposition or cathodic sputtering.

The Optical Constants of Evaporated Layers of Tin.

Eight evaporated layers of tin were examined at one angle of incidence only, 65°. The reflection coefficient at normal incidence was 0·70, and the layers were stable in air.

Fig. 8.



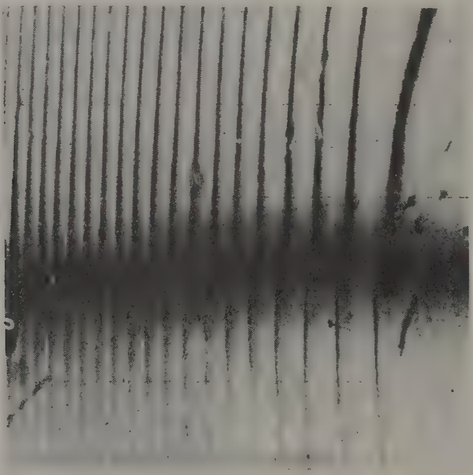
The variation of relative phase change with incidence angle for thick copper layers.

The optical constants derived for a wavelength of 5400 Å. were $n=2\cdot4\pm0\cdot2$, $k=1\cdot9\pm0\cdot1$. The accuracy of the values for the constants of copper cannot be claimed here since these measurements were made at one angle of incidence only. The optical constants for bulk tin (Hedges and Homer 1937) are given as $n=1\cdot0$, $k=4\cdot2$, and even allowing for considerable variations in the values for the bulk metal (tin is soft and so its surface would be susceptible to conditions of polish etc.) there exists considerable difference between the optical constants of the layers and of the bulk.

The constants for the bulk metal yield theoretical values for ρ and Δ of 0·82, and $0\cdot130\,\lambda$. These values were checked experimentally. Some plane surfaces of tin were prepared by allowing drops of the molten metal to fall on a hot glass plate, and to cool there slowly. When separated from the glass the tin surface had the finish of the glass plate, and also was presumably free from any distortion due to cold working. The ellipticity of the light reflected from these surfaces was measured, and the relative phase change was determined using the interferometric technique in reflection.

A photograph of the interference fringes obtained by reflection in an interferometer formed between a freshly prepared evaporated silver layer (R=0·88, T=0·06) and the cast tin surface (R=0·80) is shown in fig. 9. In order to use reflection fringes in this connection it is desirable to improve the visibility by polarizing the two halves of the field at right angles to one another and to observe the doubling along the line between the halves. The accuracy of measurement on these fringes is not equal to that obtainable with transmission fringes, but it is sufficient for the present case.

Fig. 9.



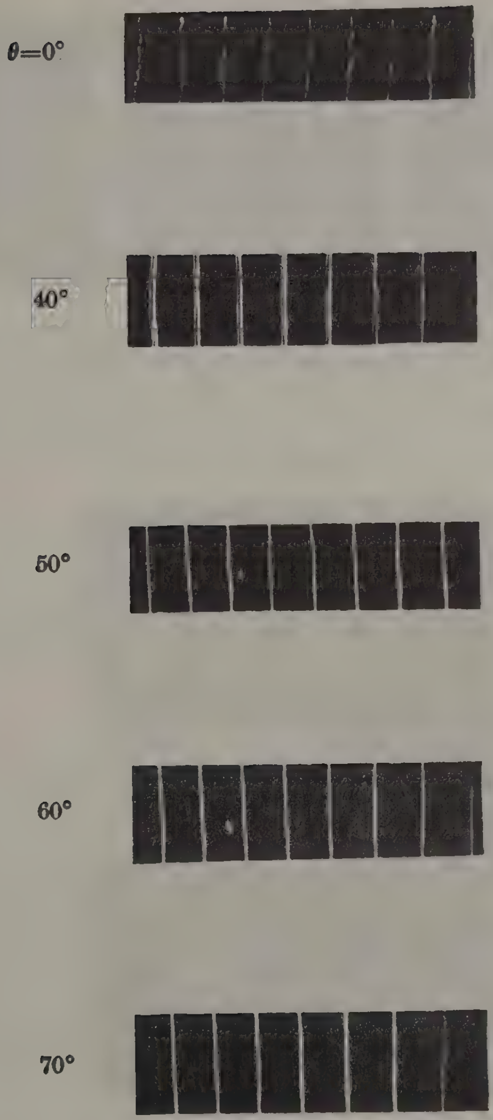
Reflection fringes at 65° incidence in the surface of a cast tin specimen.

The results of this experiment together with the experimental values for the layers, and the theoretical values for bulk tin are included in the Table below.

TABLE.

	ρ	Δ/λ	R	n	k
Theory Bulk	0·84	0·133	0·82	1·0	4·2
Exp. Bulk (cast)	0·82±0·01	0·13±0·01	0·80		
Exp. Layers	0·73±0·01	0·101±0·003	0·70	2·4	1·9

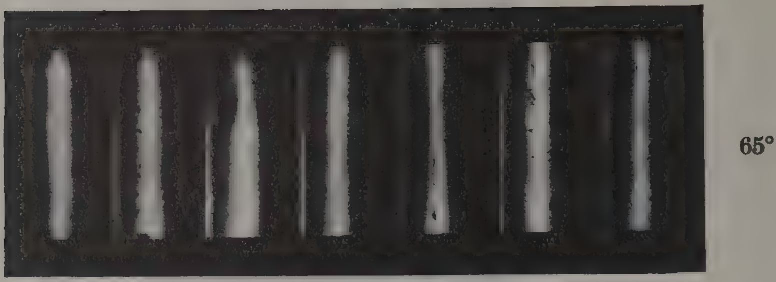
Fig. 1 (*a*).



Fringes of equal thickness in an interferometer at various incidence angles, for reflection coefficients of 90 per cent.

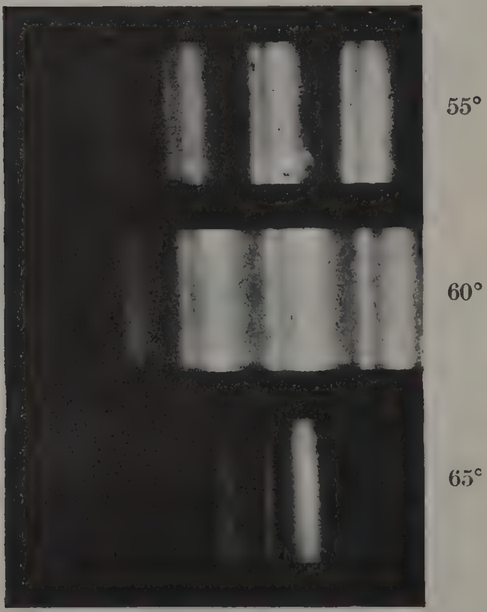
To face page 1030

FIG. 1 (*b*).



Fringes of equal thickness in an interferometer at various incidence angles, for reflection coefficients of 50 per cent.

FIG. 1 (*c*).



Fringes of equal thickness in an interferometer at various incidence angles, for reflection coefficients of 20 per cent.

The surface structure of the evaporated layers of tin would thus appear to differ from that of the metal in bulk.

Three possible explanations can be put foward to explain the discrepancy. First it may be that an oxide film was present on all the layers of tin examined, and its presence changed the optical properties of the layer. However, the layers were very stable in air, and the values of the ellipticity and relative phase change were found to be unaltered after several days, during which an oxide film might have been expected to grow and to cause further changes in the optical properties.

The second possibility (suggested by Dr. Hedges of the Tin Research Institute) is that an alloying action took place between the tin and the molybdenum filament used for the evaporation. The percentage of molybdenum in the alloy would be very small, since the filament used was of strip 0·004 in. thick, and was not visibly affected after fourteen separate evaporations from it, but the possibility remains.

A third possibility is that tin, whilst it forms a stable layer at room temperature, exhibiting specular reflection, being a low melting point layer might not form layers of a structure similar to that of the metal in bulk, even when they are opaque.

ACKNOWLEDGMENTS.

I must express my thanks to Professor S. Tolansky for suggesting these problems and for much helpful guidance throughout the work described. I am also indebted to Dr. Hedges of the Tin Research Institute for the supply of samples of very pure tin. The work was made possible by a Department of Scientific and Industrial Research maintenance grant.

REFERENCES.

ARCHARD, J., 1949, *Ph.D. Thesis* (University of London).
AVERY, D. G., 1949, *Nature, Lond.*, **163,** 916.
DRUDE, P., 1890, *Wied. Ann.*, **34,** 480.
HASS, G., 1938, *Ann. der. Phys.*, **31,** 245.
HEDGES and HOMER, 1937, *The properties of Tin*, 3rd Edition.
KHAMSAVI and DONALDSON, 1947, *Nature, Lond.*, **159,** 228.
LEVINSTEIN, 1949, *J. App. Phys.*, **20,** 306.
LOWERY, WILKINSON and SMARE, 1936, *Phil. Mag.*, **22,** 149.
O'BRYAN, H. M., 1936, *Journ. Opt. Soc. Amer.*, **26,** 122.
ROUARD, P., 1949, *Rev. d'Optique*, **28,** 569.
TOLANSKY, S., 1944, *Phil. Mag.*, **35,** 120; 1945, *Ibid.*, **36,** 225; 1948, *Multiple-beam interferometry of surfaces and films* (Clarendon Press).
WINTERBOTTOM, A. S., 1936, Private communication; 1948, *Journ. Opt. Soc. Amer.*, **38,** 12.

XCIV. *The Magnetic Deflection of Fast Charged Particles in the Photographic Emulsion**.

By C. C. Dilworth†, S. J. Goldsack‡, Y. Goldschmidt-Clermont†, and F. Levy‖.

[Received Aug. 1, 1950.]

[Plate XXXIV.]

§ I. Summary.

Electron Sensitive Nuclear Research Emulsions have been exposed at sea level, and at the Pic du Midi (2800 m.) in magnetic fields of the order of 34,000 gauss, in order to test the possibilities of curvature, within the emulsion itself, of the tracks of charged cosmic ray particles.

The existence of measurable deflections of long tracks has been established by comparison of the results of measurements on field and no field plates, and the method has been applied to determine the sign of particles in a group of cosmic ray phenomena in the exposures at high altitude.

§ II. Introduction.

In a dense medium such as the photographic emulsion, the multiple scattering of a charged particle is so high as to mask the continuous curvature produced by a magnetic field except under high fields and on long tracks.

The angle of magnetic curvature θ_m is given by

$$\theta_m = \frac{180}{\pi} \cdot \frac{3\times 10^{-4}}{p} \mathrm{H}t \text{ degrees,}$$

where H is the magnetic field in gauss, p is the momentum in MeV./c, t is the track length in cms.

The arithmetic mean angle of scattering is given by $\bar{\theta}_s = 100{\cdot}4\sqrt{t}\mathrm{G}/p\beta$ degrees where 100·4 is a constant calculated for the emulsion, β is the ratio of the velocity to that of light, and G is a slowly varying function of β and t of the order of 4. Hence the ratio of the magnetic deviation to the mean scattering angle is

$$\theta_m/\theta_s = \beta\mathrm{H}\sqrt{t}/6050\,\mathrm{G} = \gamma_0\beta\sqrt{t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

---

* Communicated by Professor P. M. S. Blackett, F.R.S.

† Centre de Physique Nucléaire, Université Libre de Bruxelles.

‡ University of Manchester.

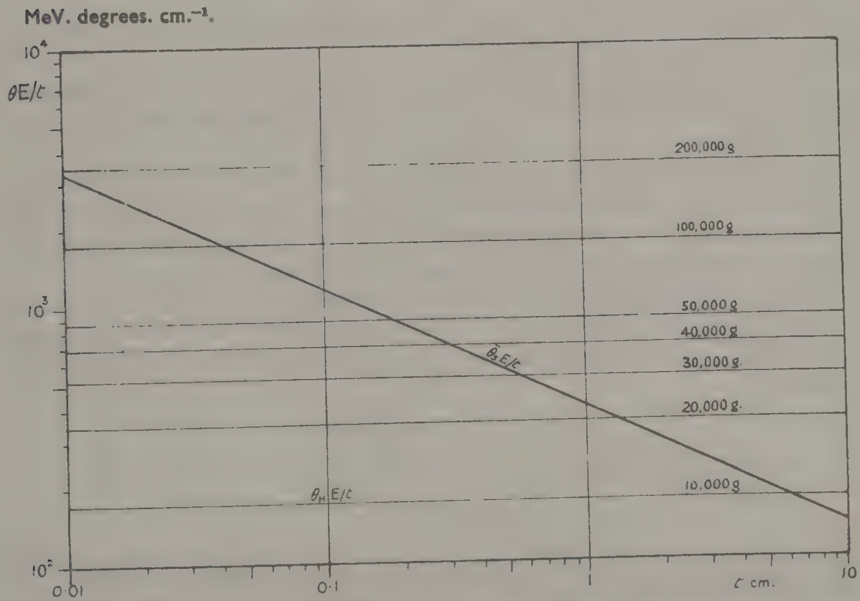
‖ University of Manchester, on leave of absence from Centre National de la Recherche Scientifique, Paris.

where γ_0 is a slowly varying function of track length and of β and is proportional to the field H *.

For a field H=34,000 gauss, and for a relativistic particle of track length 1 cm. $\gamma_0 \simeq 1\cdot46$.

In fig. 1 the sloping line shows the quantity $\bar{\theta}_s E/t = 100\cdot4\phi/\sqrt{t}$ as a function of track length. The horizontal lines show $\theta_H E/t$ for relativistic particles, for various field strengths. The figures show that for H=30,000 gauss the ratio $\theta_H/\theta_s=1$ when $t \simeq 5$ mm. This indicates the order of length of track required to separate the magnetic deflection from the scattering deflection for relativistic particles. For slower particles the length must be greater by a factor $1/\beta^2$.

Fig. 1.



Plot of $\bar{\theta}_s E/t$ as a function of track length t. The horizontal lines show $\theta_m E/t$ for relativistic particles.

In the following paragraphs we discuss how the magnetic deflection may be separated from the scattering and used to determine the sign of the particle, to improve the estimate of momentum, and to estimate the ratio of deflection to scattering $\beta\gamma_0\sqrt{t}$. Since γ_0 is nearly independent of the type of particle and of its momentum, its mean value for a number of tracks may be regarded as a test of the scattering theory, or as a calibration of the plates.

---

* This definition of $\bar{\theta}_s$ and of γ_0 differs from that of Moyal (1950) by a factor $\sqrt{\pi/2}$ since $\bar{\theta}_s$ is the root mean square scattering angle in his paper.

§ III. METHODS OF ANALYSIS.

(1) *Estimation of the Curvature.*

To measure magnetic deflection and scattering we use the customary method of dividing the track into a set of equal cells, and we measure the direction θ_i of the track in the ith cell with respect to an arbitrary origin. The deviation $\xi_i=\theta_{i+1}-\theta_i$ between two successive tangents to the track is the sum of a constant magnetic deflection and a random scattering deflection with zero mean.

Except at highest momenta we may neglect measurement errors for the present discussion. If we assume the distribution of scattering to follow a Gaussian law, the best estimate for the curvature is the sum of this set of independent measured deviations $\sum_1^n \xi_i$.

This sum is equal to the total deviation between tangents at the two ends of the track* $\phi=(\theta_n-\theta_0)$.

This result is confirmed by a more refined analysis, taking into account small measurement errors (Moyal 1950) or using sagitta measurements (Scott 1949).

The actual distribution of scattering is closely similar to a Gaussian distribution, but has a rather large "single scattering tail". The effects of this tail are most simply eliminated by subtracting large deviations from the total deflection in estimating the curvature.

The standard error on the estimate of the magnetic deflection θ_m from the total deviation ϕ is the sum of the independent errors, the scattering error $(\frac{1}{2}\pi)^{\frac{1}{2}}\bar{\theta}_s$ and the measurement error s_0 on each of the first and last settings.

$$i.\,e.\ \theta_m=\phi\pm(\tfrac{1}{2}\pi\bar{\theta}_s^2+2s_0^2)^{\frac{1}{2}}.$$

In most cases $2s_0^2 \ll \frac{1}{2}\pi\bar{\theta}_s^2$, so we may write

$$\theta_m=\phi\pm(\tfrac{1}{2}\pi)^{\frac{1}{2}}\bar{\theta}_s.$$

The relative error on the determination of the momentum p from the magnetic deflection is therefore approximately

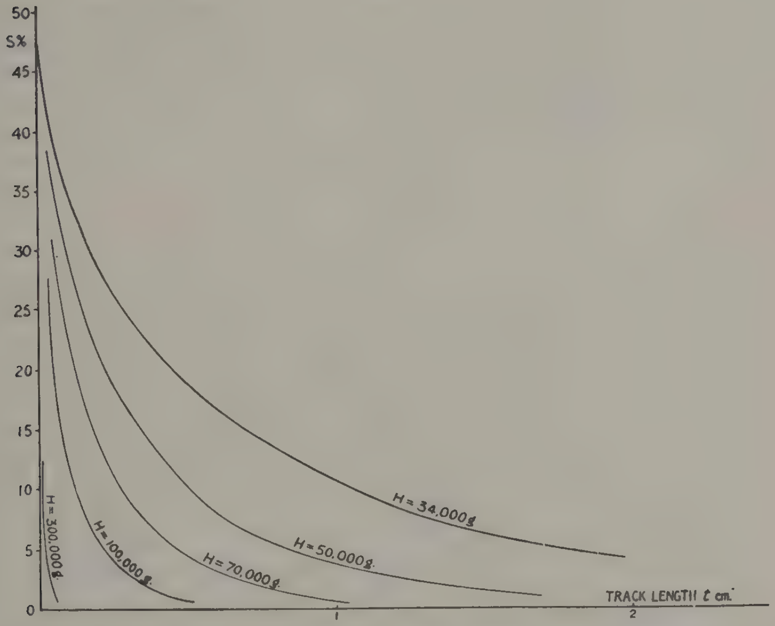
$$\frac{\epsilon(p)}{p}=\frac{\epsilon(\theta_m)}{\theta_m}=\sqrt{\frac{\pi}{2}}\cdot 1/\beta\gamma_0\sqrt{t}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

[*] This result depends on the statistical independence of the successive deflections ξ_i. If the track were a perfect arc, and measurement errors were the only cause of scattering, the actual observations θ_i, and not the deflections ξ_i would be independent. The best estimate for the curvature would then be given by the mean value of the deflection over half track distances, $\frac{2}{n}\sum_1^{n/2}\left(\theta_{\frac{n}{2}+i}-\theta_i\right)$ since this statistic involves each of the n independent observations with equal weight. These conditions are encountered in cloud-chamber measurements on particles of high momentum, and may exist for such particles in the plate if higher fields become available.

The errors on the determination of momentum from scattering are discussed in paragraph (5). The variation of these errors with energy of the particle is shown in fig. 7, where $\frac{\epsilon(p\beta)}{p\beta}\sqrt{t}$ is plotted against $p\beta$ for typical values of the measurement errors. The horizontal lines show the relative error from magnetic curvature, together with the form of the curves for protons and π mesons, allowing for their becoming non relativistic.

Fig. 2.



The proportion S per cent of particles of track length t, which, owing to multiple scattering, will give a total deflection in the wrong sense, as a function of t for various field strengths.

It can be seen that over wide ranges of energy the momentum is determined with greater precision from scattering observations than from magnetic deflection in moderate fields. However, for very high fields (of the order of 100,000 gauss), and for high momentum particles, the estimate from magnetic deflection is more accurate.

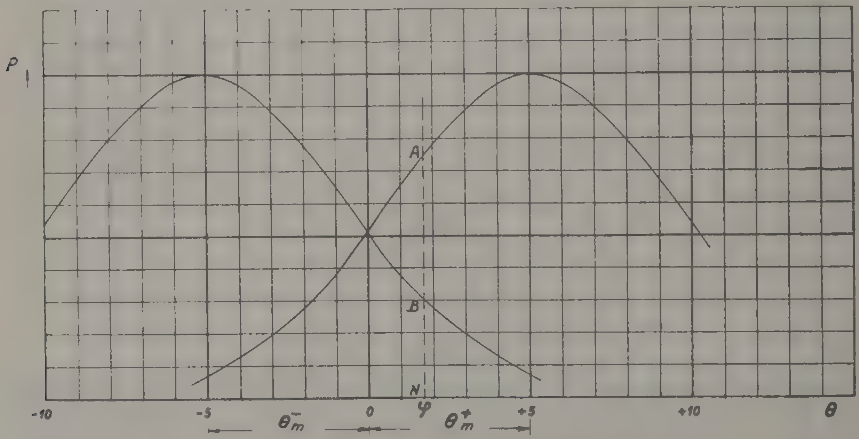
(2) *Determination of the sign of particles.*

In many tracks from cosmic-ray events the direction of motion of the particle is known. The determination of its sign then depends simply on the sense of its deflection in the magnetic field. Owing to the superposed scattering there exists a certain probability that a particle will have a total deflection in the wrong sense. This probability is a function of track length and of magnetic field, and is plotted in fig. 2. Here it will be noticed that for a track of 1 cm. and a field of 34,000 gauss, the probability is 12 per cent.

If, however, the momentum can be determined with fair accuracy from scattering measurements, or from other considerations, we may frequently assess the sign with greater accuracy. For consider fig. 3. Here are shown two distributions representing the probability distribution for the total deflection ϕ if the particle has a positive or negative sign respectively, the mean value θ_H being predetermined from the knowledge of the momentum. If we observe for some particular track, a total deviation ϕ, then clearly the ratio of the probabilities of the sign being positive or negative is given by the ratio of the ordinates of the two distributions at the position ϕ, *i. e.*

$$\frac{P(+|\phi)}{P(-|\phi)}=AN/BN.$$

Fig. 3.



The probability distributions for the total deviation for a positive or negative particle whose mean is known from scattering measurements to be θ_H.

This assumes that a priori there are equal chances of the particle having either sign. A more rigorous analysis of the situation is given by Moyal. With $p(+)/p(-)$ the ratio of the a priori sign probabilities, his conclusion is expressed by

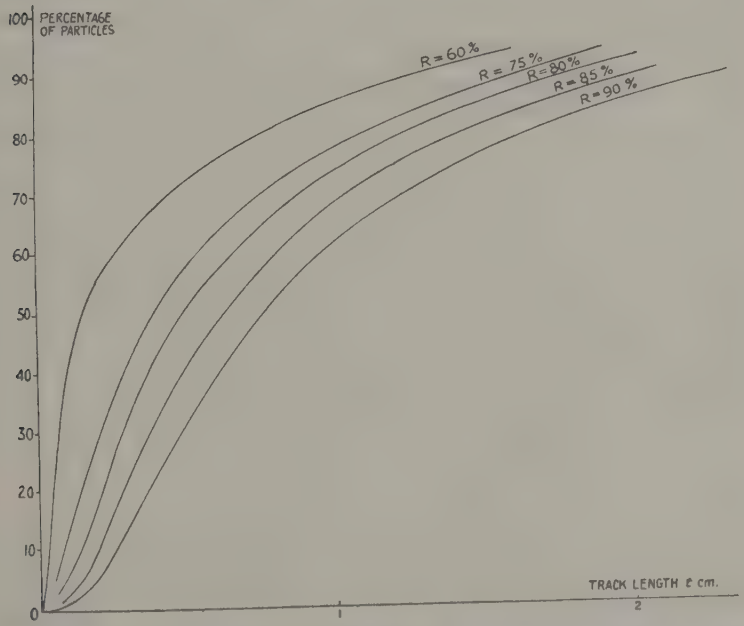
$$P(+|\phi)=\left\{1+\frac{p(+)}{p(-)}\exp-\frac{4\beta^2}{\pi}\gamma_0\gamma t\right\}^{-1}, \quad . \quad . \quad . \quad . \quad (3)$$

where $\gamma=\phi/\bar{\theta}_s\sqrt{t}$ is the estimated value of γ_0.

In the present experiment we assume the ratio $p(+)/p(-)$ to be always unity. Fig. 4 shows the proportion of positive particles of length t which will yield a positive sign with probability $P(+|\phi)$ greater than the fixed values marked on the curves for a field of 34,000 gauss.

It should be pointed out that in the event of a track occurring during the inevitable "no-field time" and being believed to be a field track, one will, in general, attribute a sign with a probability greater than 50 per cent. The fraction of no field tracks which would give sign probabilities greater than the same fixed values as in fig. 4 is shown in fig. 5. It will be seen that this fraction can be rather high for long tracks. For this reason it is important to reduce to a minimum the no-field time, and to keep the plates horizontal while out of the field.

Fig. 4.



The proportion of particles, of track length t, which will give the correct sign with probabilities greater than R per cent for $H=34{,}000$ gauss and for various values of R, as a function of t.

(3) *Estimation of Scattering.*

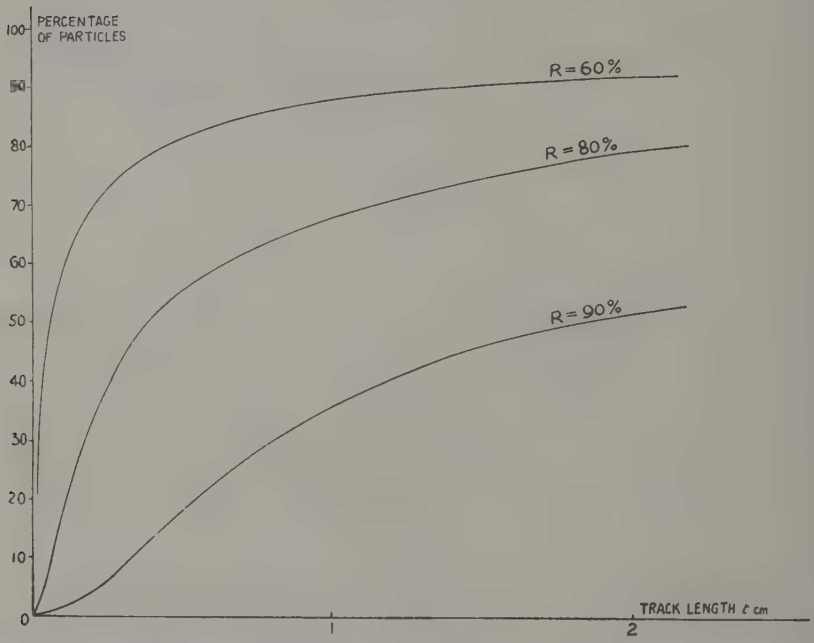
To estimate the multiple scattering, the direction of the mean line through the grains in each cell is carefully measured, using an accurate eyepiece goniometer, fitted with a suitable graticule*.

---

* The method of measurement of sagitta, described recently by Fowler (1950), could also be used for these measurements. We are now investigating the advantages of the two methods.

Two methods are available for the calculation of the momentum from the measured angles. Either the arithmetic mean deviation, or the root mean square is calculated. In either case the effects of the single scattering tail are eliminated by applying suitable weighting functions, as for example, for the mean deviation, a sharp cut off of all angles greater than four times the mean. Such a sharp cut off can be shown to introduce a much greater uncertainty in the root mean square than in the mean deviation.

Fig. 5.



The proportion of particles of track length *t* which, in the absence of a field, may be expected to yield a sign probability (correct or incorrect) greater than R per cent for various values of R. The proportion which will yield the incorrect sign with these probabilities is, of course, just one-half those given by these curves.

(4) *Determination of the Errors.*

Errors of two types occur in the measurements of multiple scattering and of the total deflection.

(*a*) Distortion of the emulsion and inaccuracy of the microscope stage.

(*b*) Measurement errors, which we call spurious scattering or noise; statistical errors, due to the finite number of cells available; and an uncertain amount of " smoothing", a systematic reduction in scattering angles due to the treatment of the track as a straight line within each cell.

The first type of error is more serious for determination of the magnetic deflection, the second in the determination of multiple scattering. The emulsion, especially when wet, is easily distorted. Even slight distortions could limit the measurements to particles of low momentum. Their effects can be reduced by using thin emulsions, but only at the expense of the frequency of long tracks. Thin plates were used for the sea level exposure, but recent advances in the processing of thick emulsions (Dilworth *et al.* 1948 and 1950) and in the measurement of distortion (Cosyns *et al.* 1950) made possible the use of thicker emulsions for some of the later experiments. Measurements of distortion on both sets of plates are described in § IV., where it is shown that they have been reduced to a satisfactory level. For the measurements on high energy tracks the movement of the microscope stage is controlled throughout the measurements by an optical method. Both Cooke and Baker microscopes show a deflection less than 0·4′ over a displacement of 20 mm.

The determination of the second type of error depends on the method of analysis used.

In using the root mean square determination of scattering, systematic effects of smoothing and of spurious scattering may be eliminated by the method suggested by Moyal. He points out that true and spurious scattering have a different statistical nature. True scattering yields a set of independent deflections ξ_i while the presence of measurement errors on each setting θ_i is shown to introduce a negative correlation between the successive ξ_i's. On the other hand, smoothing introduces a positive correlation of a similar nature. Both effects are eliminated on the average by using as an estimate for the scattering

$$\tfrac{1}{2}\pi\bar{\theta}_s^2=(\Sigma\xi_i^2-\phi^2/n)+2(\Sigma\xi_i\xi_{i+1}-\phi^2/n). \quad . \quad . \quad . \quad . \quad (4)$$

The terms containing the total deviation ϕ are small corrections to allow for the mean deflection, and the second bracket term is a correction to the sum of squares to allow for the combined effects of noise and smoothing.

It is shown by Moyal that neglecting the departure of the scattering law from the Gaussian Law, the relative error on this estimate of $\bar{\theta}_s^2$ is given by

$$\epsilon(\bar{\theta}_s^2)/\bar{\theta}_s^2=(2/n)^{\frac{1}{2}}\left\{\left(1+\frac{4w}{\pi\bar{\theta}_s^2}\right)^2+2\right\}^{\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where w is the mean value of $(\Sigma\xi_i\xi_{i+1}-\phi^2/n)$.

Alternatively, using the arithmetic mean as an estimate for $\bar{\theta}_s$ the systematic effects of noise and smoothing are eliminated by the methods described by Goldschmidt-Clermont *et al.* (1948).

Smoothing is taken into account by using the angle between cells separated by one cell distance, *i. e.* $(\theta_{i+2}-\theta_i)=(\xi_i+\xi_{i+1})$, and then increasing the measured scattering by 4 ± 2 per cent *. The noise level for the track is not known, but is estimated from measurements on similar tracks of such high momentum that the true scattering is negligible, and the observed scattering is reduced accordingly.

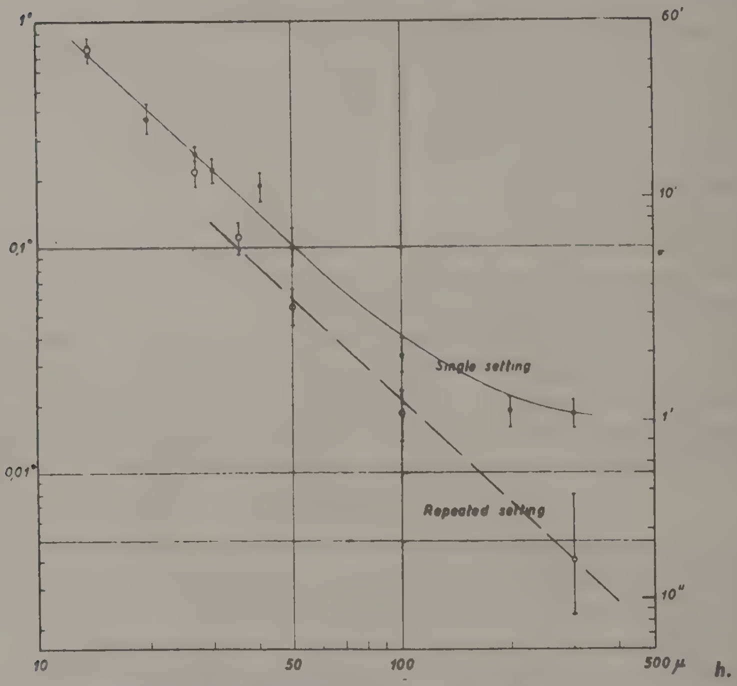
---

* If this method were applied to the root mean square estimate we would have to calculate $\frac{1}{2}\Sigma(\xi_i+\xi_{i+1})^2=\Sigma\xi_i^2+\Sigma\xi_i\xi_{i+1}$. The second term is just half the correction suggested by Moyal.

The $(n-1)$ measured angles are not statistically independent. This effect is allowed for by treating $\frac{2}{3}$ of the observations as independent. Corrected for the effect of cutting off large angles, the statistical error is

$$0{\cdot}54\,[\tfrac{2}{3}(n-1)]^{-\frac{1}{2}}.$$

Fig. 6.



The experimentally observed spurious scattering per cell as a function of cell length for single settings and for multiple settings.

(5) *Choice of Cell Length.*

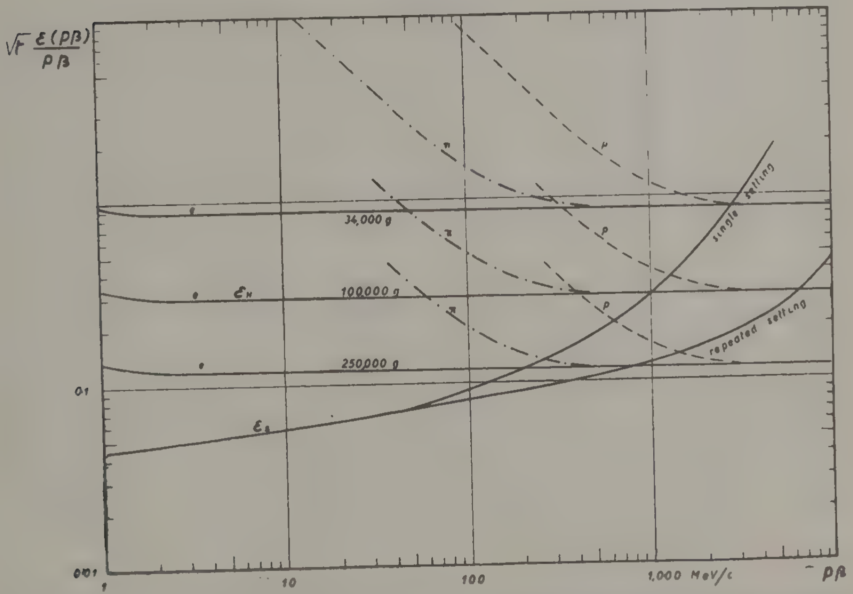
The choice of cell length results from a compromise between two conflicting requirements: a short cell reduces the statistical errors while a long cell decreases the noise level. The optimum value is that which renders minimum the total error and can be calculated from the empirical law of variation of noise level with cell length. Measurements have been made on high energy tracks to determine the magnitude of the spurious scattering and its approximate variation with cell length.

Analysis of the experimental results shows that the noise may be divided in two parts:

(1) One component, which we may call setting error, is introduced by the inability of the observer to set the line of the graticule accurately along the mean line of the grains. This error can be determined by repeated settings on a cell, and may be reduced to as low a level as is required by averaging a sufficiently large number of settings.

(2) The second and more fundamental source of spurious scattering is the spread of the grains about the true direction of the track. The mean angle $\bar{\epsilon}$ between the track and the best mean line which can be drawn through the grains can be calculated by supposing the grains to lie at random on either side of the track and at a mean distance r from it. It is found that

Fig. 7.



The relative error of estimation of energy at optimum cell length from scattering using single and multiple settings as a function of energy. The auxiliary lines show the error in the estimate of momentum from magnetic deflection for electrons, protons and π mesons, given by equation (2).

$$\bar{\epsilon} = \frac{Ar}{\sqrt{d}}\,h^{-3/2}.$$

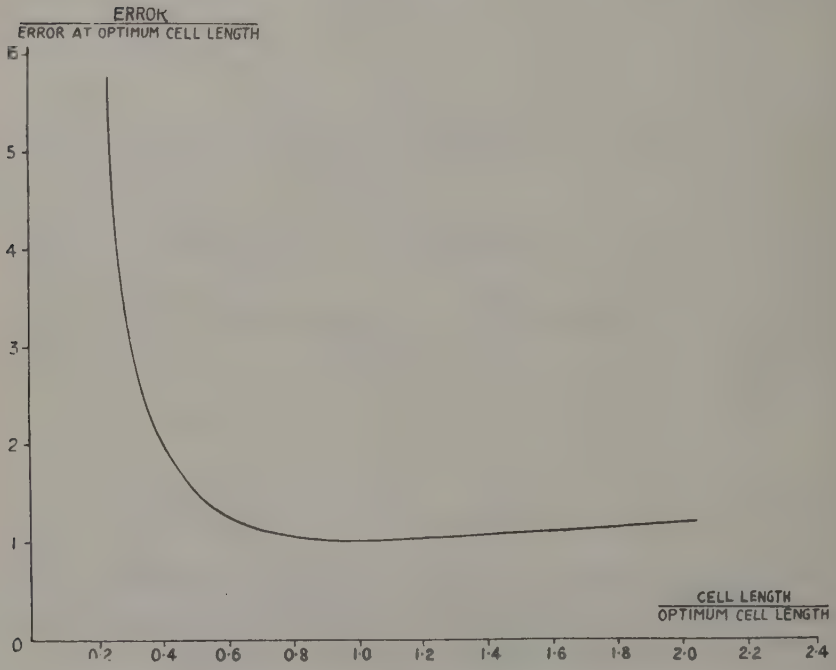
Where d is the grain density, and h the cell length; r may be assumed to be of the order of the radius of a grain, and $A \simeq 1$.

Fig. 6 shows the observed variation of noise level with cell length. The upper curve gives the spurious scattering for single settings, the lower one that for averages of several settings. The number of settings to be averaged increases as the cell length increases. The lower curve may be assumed to represent the noise level arising from the second source. It will be seen that it is less than 1′ for cell length of 200 μ.

The value of the optimum cell length depends on the type of analysis used. With Moyal's method, where the noise level is calculated from the set of observations, it can be shown that the minimum error is obtained when the total mean square scattering is the sum of two nearly equal terms due to noise and true scattering

$$\frac{1}{n}\cdot\frac{1}{2}\cdot\pi\bar{\theta}_s^2 = s_0^2/\rho,$$

Fig. 8.



The variation of total error of momentum estimate by scattering with cell length.

where ρ is a constant of the order of 0·5. The values calculated for the total error for a single setting and for multiple settings are presented in fig. 7 as a function of momentum.

Fig. 8 shows the theoretical variation with cell length of total errors. Owing to the rapid variation of spurious scattering with cell length the total error is seriously increased by using too short a cell length. On the contrary, the use of somewhat too large a cell length will increase it only slightly. It will be seen from the figure that the use of a cell length equal to twice optimum increases the error by only 20 per cent.

For the alternative method of analysis it is necessary to use a somewhat greater cell length, since the effect of setting errors is estimated from measurements on other tracks, and we require that the uncertainty introduced should be small. As shown, however, this does not seriously increase the total error.

§IV. EXPERIMENTAL METHODS.

(1) *Exposure and Processing of the Plates.*

(A) *Sea level plates.*

The photographic plates were exposed in a vertical position in an electromagnet being built for the University of Colombo, Ceylon, at the firm of Metropolitan Vickers Ltd., Manchester. The electromagnet was similar in design to that described by Blackett (1934) for use in cloud chamber studies of cosmic ray events, having poles of 10 in. diameter. For the present experiments the poles were coned down at an angle of 60° with good quality high permeability steel (kindly supplied by Metropolitan Vickers Ltd.) to a diameter of 1 in. The gap used was 1 cm. With this arrangement and a power consumption of 6 kilowatts the field was 34,000 gauss.

The plates were Kodak NT4 plates 200 μ thick emulsions and 2 in. × 2 in. dimensions. They were carried in a horizontal position from London as soon as they were dry, and two were put in the magnetic field. Two other plates were kept in a vertical position in the Physical Laboratories, Manchester University, as controls. Both batches of plates were exposed for 20 days and then developed together. The method of " Temperature Development" was used, the developer being Azol. Great care was taken in treatment of the emulsions, especially in the drying stages, to avoid distortion. The plates were processed in a horizontal position.

(B) *Pic du Midi.*

These plates were exposed, also in a vertical position, in the field of the Blackett magnet which has recently been transported to the Pic du Midi (2800 m.). The pole pieces were 7½ cm. diameter, made of soft iron with permendur tips and gap about 1 cm. With a power consumption of 12 kilowatts the field was 34,000 gauss. Ilford G.5 plates, 200, 400 and 600 microns thick and 5 inches square were used.

The large area was taken to make the maximum use of the field and also to reduce distortion due to edge effects. The plates were dried horizontally at Ilford Ltd., to avoid stresses being introduced into the emulsion, which can later give rise to distortion. They were brought to Pic du Midi in a horizontal position to reduce the number of long no-field tracks, exposed for three weeks, and then brought back for development, again horizontally, and at low temperature to reduce fading.

Development was by Temperature Development, using amidol at low pH (Dilworth *et al* 1950). Great care was taken to avoid sudden changes of temperature, and of pH. Fixing was carried out at low temperature

(10 to 14° C.), with as little agitation as possible consistent with efficient fixing. The plates were washed in a slow non-turbulent stream of water. During the drying the plates were surrounded with a "guard ring" of wet gelatine plates; the whole placed on the glass bottom of an enclosed box, in which the humidity was regulated by the use of a "chimney". The plates were gently warmed by an infra-red lamp placed at some distance below the box. In this way the plates were maintained as far as possible with equal conditions of humidity all over the surface.

(2) *Measurements on the Tracks.*

(i) *Scanning.*

Sea level plates.—The plates were carefully scanned on a Cooke M4000 binocular microscope fitted with a Cooke 45× fluorite oil immersion objective and a pair of 8× compensating eyepieces. Many long straight tracks, mostly due to cosmic-ray mesons, were found. Some 30 tracks of approximately 1 cm. in length were found within the area covered by the high magnetic field, and a similar number of long tracks were found in the control plates.

There were fewer low-energy tracks on the no-field than on the field plates, probably because the former were kept inside a building with thick walls and floors.

Pic du Midi plates.—These plates were scanned on both Cooke M4000 and Baker 4BQ microscopes, using either Cooke 20× dry objective or Cooke 45× fluorite oil objective, and a pair of either 10× or 8× compensating eyepieces. Among the events noted those were chosen for measurement which possessed tracks of sufficient length. In 30 cm.^{2} of one 600 μ plate 40 tracks were found suitable for measurement. This does not include tracks of electron pairs, which are only found consistently with more careful scanning.

(ii) *Measurements.*

The method of measurement of scattering and of deflection has been mentioned in § III., and is described elsewhere (Y. Goldschmidt-Clermont 1950).

To define the direction of the track accurately we require a very fine straight line. Following the procedure of Braddick (1950) we have produced graticules consisting of one line, or of two parallel lines about 1 mm. apart, by ruling with a fine point on a photographic plate. Each line is of width about 0·01 mm. These have proved very suitable for this work. The grains are set symmetrically along the single line or down the space between the two lines. Different cell lengths are obtained by using eyepiece stops, and changing the magnification.

On each plate the distortion has been measured by the method of Cosyns and Vanderhaeghe, using measurements on steeply dipping tracks, and corrections applied to each track according to its dip and its orientation with respect to the distortion vector. As an additional check in the case of the sea level plates no-field control plates have been examined.

On these plates 16 tracks are deflected by less than 0·1° over distances of the order of 1 cm. One track in this group ran for 5 cm., and disregarding the 0·5 cm. near each end where it is in the danger zone near the edge of the plate, the deflection was 0·05°. Corresponding figures for the high energy particles in the field plates are that eleven such tracks were deflected by less than 0·1°.

Direct measurements have been made on the Pic du Midi plates. The distortion is, of course, least on long tracks, corresponding to their small dip. As examples of the total distortion over a whole track length ∼1 cm., we found 2′ in one case and 25′ in another.

In many cases it has been possible to identify a particle by comparison of observations on the grain density and scattering. (Camerini *et al.* 1949, Cosyns 1949, Fowler 1950, Goldschmidt-Clermont 1950, Occhialini 1949).

§V. EXPERIMENTAL RESULTS.

(1) *Determination of the Curvature.*

Eleven tracks on the field plates and eight tracks on the no-field control plates have been examined and analysed by Moyal's method. The results for these tracks are presented in Tables I. and II. respectively. In figs. 9 and 10 are presented detailed plots of the tracks.

The values of the ratio

$$\gamma=\phi/\bar{\theta}_s\sqrt{t}$$

have been calculated for both types of track. For no-field tracks these should be distributed normally about zero mean. The standard deviation of the distribution is given by

$$\epsilon(\gamma)\doteqdot\sqrt{\frac{\pi}{2t}}\left\{1+\frac{2}{\pi}\gamma_0^2 t\,\frac{\epsilon^2(\bar{\theta}_s)}{\bar{\theta}_s^2}\right\}^{\frac{1}{2}},$$

the second term in the bracket is frequently small.

For field tracks it should have mean value γ_0 and the same standard deviation.

The weighted mean value of $|\gamma|$ obtained from the no-field tracks is

$$|\bar{\gamma}|=0{\cdot}49\pm0{\cdot}39.$$

For the field tracks the value $|\gamma|$ is found to be

$$|\bar{\gamma}|=1{\cdot}96\pm0{\cdot}44.$$

The difference is 1·47±0·59.

It seems that the presence of a significant effect from the magnetic field is established. To show clearly the effect we have plotted in fig. 11 the value of $g=\gamma\sqrt{t}$ which is subject to a standard error approximately $\pm\sqrt{(\pi\ 2)}$ against the square root of track length $\sqrt{t}$, field and no field tracks being plotted on the same diagram.

The mean value of $|\bar{\gamma}|$ for the field tracks is consistent with the calculated theoretical value $\gamma_0=1{\cdot}46$ within the rather wide errors. Many more tracks will have to be examined before it will be possible to give an experimental calibration of the scattering by this method,

TABLE I.

Field Tracks on Sea-Level Plates.

Track	Length t cm.	Cell length h microns	Estimated optimum cell h microns	Total devn. ϕ°	Mean scattering angle $\bar{\theta}_s{}^\circ$	Momentum eV/c		$\gamma=\phi/\bar{\theta}_s\sqrt{t}$	Sign probability %
						Scattering estimate	Magnetic defn. estimate		
1.	0·83	26	11	45·5	19·4±1·2	(1·8±0·1)×10^7	(1·1±0·5)×10^7	2·53±1·4	98
2.	0·80	104	50	5·3	4·1±0·5	(8·4±0·5)×10^7	(8·7±8·6)×10^7	1·45±1·43	89
3.	0·44	104	110	0·75	0·64±0·2	(3·6±0·9)×10^8	(3·4±3·6)×10^8	1·8±2·0	82
4.	0·29	104	110	0·84	0·45±0·12	(2·8±0·9)×10^8	(0·8±0·3)×10^8	5·8±2·0	96
5.	1·79	240	160	2·44	0·88±0·15	(5·8±1·0)×10^8	(4·3±2·6)×10^8	1·87±1·0	99
6.	1·24	185	180	0·33	0·74±0·12	(5·7±1·4)×10^8	(22±44)×10^8	0·40±1·1	60
7.	1·25	240	220	0·84	0·49±0·10	(8·3±1·7)×10^8	(7·9±5·8)×10^8	1·58±1·2	96
8.	0·72	240	300	0·68	0·29±0·08	(1·1±0·3)×10^9	(0·62±0·31)×10^9	2·6±1·7	97
9.	1·30	240	350	1·53	0·34±0·14	(1·5±0·6)×10^9	(0·4±0·4)×10^9	3·4±2·1	94
10.	1·66	240	380	0·51	0·29±0·11	(1·7±0·6)×10^9	(1·9±1·2)×10^9	1·6±1·2	97
11.	2·35	240	560	1·04	0·21±0·17	(2·7±1·1)×10^9	(1·3±0·7)×10^9	3·0±2·6	80

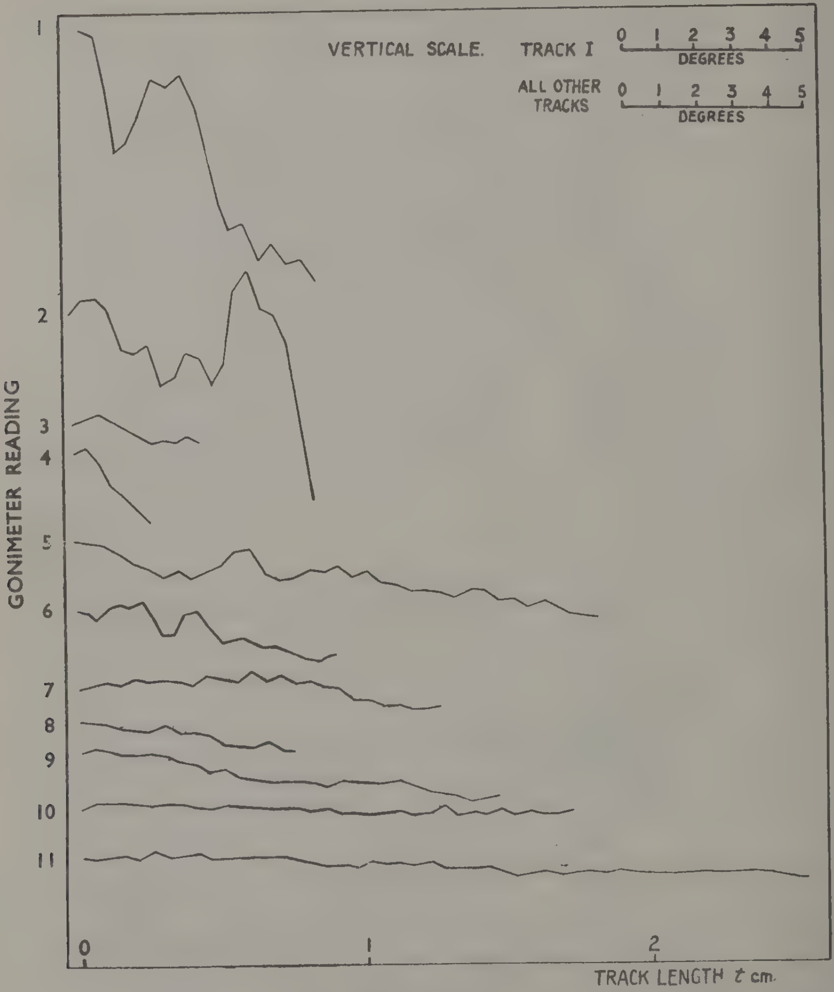
TABLE II.

No-Field Tracks on Sea-Level Plates.

Track	Length t cm.	Cell length h microns	Estimated optimum cell h microns	Total deviation ϕ°	Scattering deviation $\bar{\theta}_s^\circ$	Momentum eV/c scattering estimate	$\gamma=\phi/\bar{\theta}_s\sqrt{t}$
1.	1·89	104	100	1·56	$1{\cdot}6\pm0{\cdot}2$	$(3{\cdot}3\pm0{\cdot}4)\times10^8$	$0{\cdot}73\pm1{\cdot}15$
2.	2·25	240	170	0·83	$0{\cdot}85\pm0{\cdot}11$	$(6{\cdot}7\pm0{\cdot}9)\times10^8$	$0{\cdot}63\pm0{\cdot}84$
3.	1·08	180	190	0·56	$0{\cdot}52\pm0{\cdot}08$	$(7{\cdot}8\pm1{\cdot}2)\times10^8$	$1{\cdot}05\pm1{\cdot}25$
4.	0·79	240	290	0·16	$0{\cdot}26\pm0{\cdot}13$	$(1{\cdot}3\pm0{\cdot}3)\times10^9$	$0{\cdot}67\pm1{\cdot}68$
5.	1·30	240	320	0·03	$0{\cdot}31\pm0{\cdot}13$	$(1{\cdot}4\pm0;3)\times10^9$	$0{\cdot}09\pm1{\cdot}1$
6.	1·68	240	350	0·15	$0{\cdot}30\pm0{\cdot}06$	$(1{\cdot}7\pm0{\cdot}4)\times10^9$	$0{\cdot}40\pm{\cdot}94$
7.	1·07	180	390	0·04	$0{\cdot}22\pm0{\cdot}12$	$(1{\cdot}9\pm1{\cdot}0)\times10^9$	$0{\cdot}17\pm1{\cdot}3$
8.	2·50	240	500	0·07	$0{\cdot}19\pm0{\cdot}25$	$(3{\cdot}0\pm4{\cdot}4)\times10^9$	$0{\cdot}24\pm0{\cdot}84$

The correct sign probability has been calculated for each track from equation (4), and is also presented in Table I. It will be noticed that owing to the generally high values of γ observed, leading to the mean somewhat above $\gamma_0=1{\cdot}46$, these probabilities are quite high.

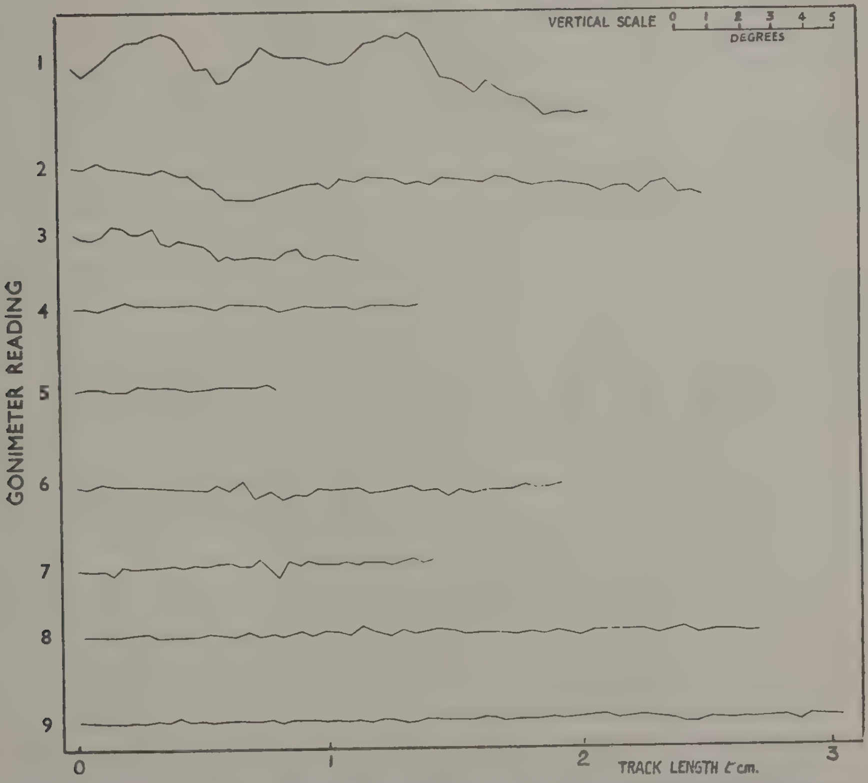
Fig. 9.



Detailed plot of eleven field tracks arranged in order of their momenta, and corresponding to the eleven tracks in Table I. Tracks 3 and 4 are two halves of the same track. For clarity negative deflections have been inverted to prevent crossing of the lines.

A discussion of the actual significance of the observed deviation is a little difficult, for clearly both the deflected and the undeflected tracks have mean deflection zero if we disregard the possible positive (or negative) excess. On the other hand, a presentation in terms of the modulus of the values of γ does not give an average which has any real meaning since the distribution about the mean is a function of track length. One way of presenting the data which seems reasonable is to calculate for each track the quantity $\{\gamma/\epsilon(\gamma)\}$. For the no-field tracks this should be distributed about zero with unit standard deviation. It follows that the sum of squares of terms such as $\Sigma\{\gamma/\epsilon(\gamma)\}^2$ should be distributed according to the so-called χ^2 distribution, and the probability of the observed value being exceeded is easily found from tables of this distribution. The same analysis applied to the field tracks will then show whether the observed set of values of γ are consistent with the hypothesis that they belong to an undeflected (no-field) distribution. For the eleven field tracks the value was

Fig. 10.



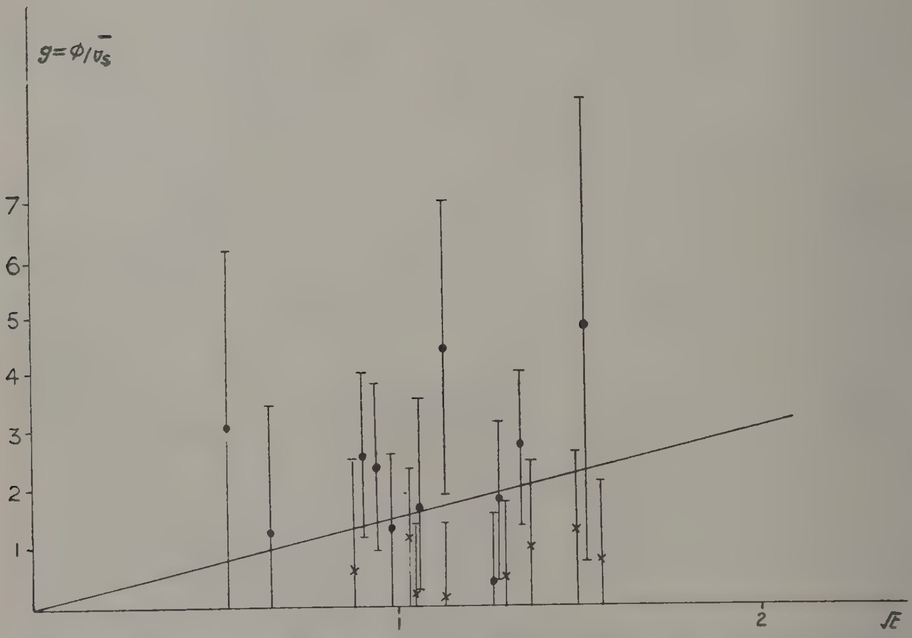
Detailed plot of nine no-field tracks. The first eight are those listed in Table II., the remaining one being a very high energy track, which is included to show the low level of distortion.

$$\Sigma\{\gamma/\epsilon(\gamma)\}^2=22\cdot5$$

and reference to tables shows that this value is exceeded with a probability of only 3 per cent in a random sample from a normal distribution with zero mean. It follows that the observed deviations are not consistent with the hypothesis that they represent merely a fluctuation from a set of undeflected tracks.

Moyal's statistic can give a negative value for the true scattering where the total relative error is large. This happened for some of the no-field tracks since most of them were of rather high energy, and these were discarded. Tracks of similar energy for which the calculation gave

Fig 11.



The observed ratio $g=\phi/\bar{\theta}_s$, of the total deviation ϕ to the mean scattering deviation $\bar{\theta}_s$, for eleven field tracks ● and eight no-field tracks × as a function of $\sqrt{t}$. The straight line is the line $g=\gamma_0\sqrt{t}$ for $\gamma_0=1{\cdot}46$.

a positive true scattering would be accepted. This introduces a bias tending to underestimate the average energy of the tracks. When the test is applied we are, therefore, not surprised to find that the values of γ observed are significantly small. The value obtained for the eight no-field tracks was $\chi^2=2{\cdot}16$. For a random sample there is a probability of 7 per cent that χ^2 should be smaller than this. Operation of a similar bias in the case of field tracks could only have reduced the calculated significance, so that our evaluation of it is a lower limit.

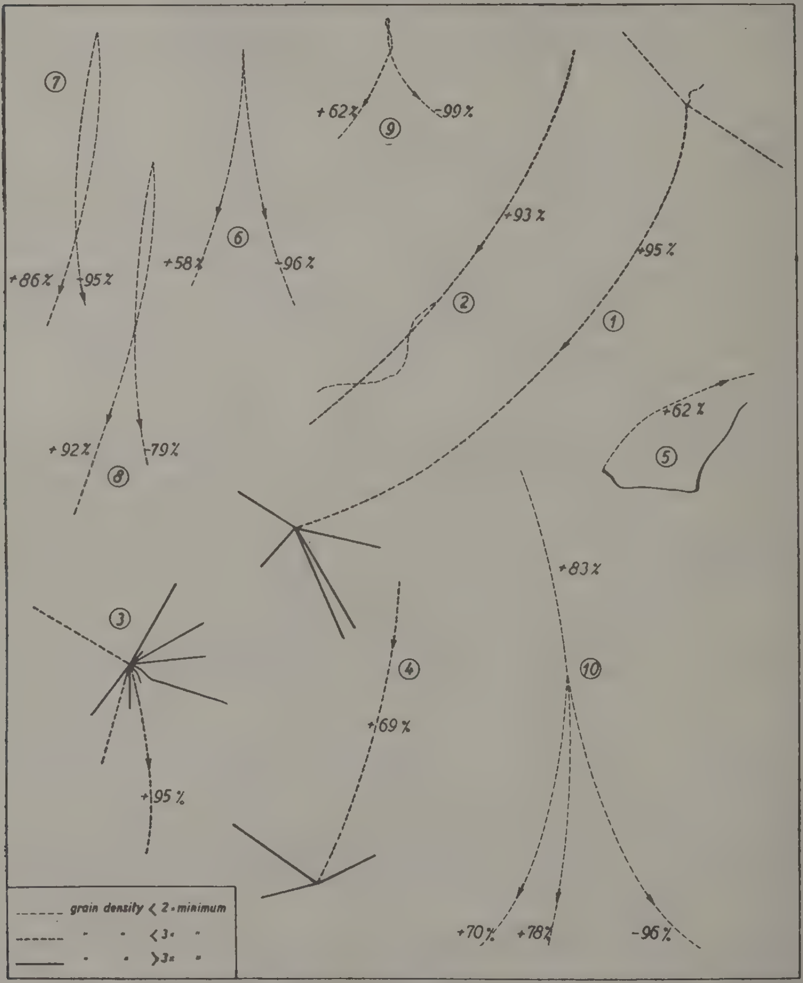
TABLE III. Cosmic-ray Events.

Event	Track length t cm.	Scattering angle°/cm. $\bar{\theta}_s/\sqrt{t}$	$\phi°$ Total deviation	Energy MeV.	Grain density grains/100 μ	Nature of particle	β	γ	Sign	Probability %	
1. Track linking double star	2	1·0	2·28	185±17	57±3	Proton	0·6	2·0	+	95	
2. Track giving rise to knock-on electron	1·3	0·75	1·32	255±85	61±3	Proton	0·6	2·2	+	93	
3. Track leaving star	1·5	1·34	4·36	140±10	70±4*	Proton	0·5	4·4	+	95	* These tracks are from a different plate with a different grain-density—ionization curve.
4. Primary of star ..	1·5	1·1	0·75	170±17	60±2	Proton	0·5	0·84	+	69	
5. π-μ-e decay electron	0·23	11	6·5	27±3	—	Electron	$\sim$1	2·5	+	62	
6. γ-ray pair .. (*a*)	0·7	4·25	4·38	70±7	36±2	Electron	$\sim$1	1·48	−	86	
(*b*)	0·9	1·49	2·76	220±18	36±2	Electron	$\sim$1	2·06	+	95	
7. γ-ray pair .. (*a*)	1·1	1·48	4·67	220±17	—	Electron	$\sim$1	2·77	−	96	
(*b*)	0·9	10·03	1·73	29±3	—	Electron	$\sim$1	1·86	+	58	
8. γ-ray pair .. (*a*)	0·6	5·9	1·85	50±7	36±2	Electron	$\sim$1	0·52	−	79	
(*b*)	0·5	4·0	5·32	75±10	36±2	Electron	$\sim$1	2·68	+	92	
9. γ-ray pair .. (*a*)	0·94	10·6	21·7	28±3	—	Electron	$\sim$1	2·18	−	99	
(*b*)	0·4	15·3	3·9	18±2	—	Electron	$\sim$1	0·64	+	62	
10. Giraffe primary (*a*)	1·0	0·11	0·07	3, 400±500†	—	?		0·84	+	83	† These tracks are of an unknown nature. The calculations have been done assuming they are electrons. If they are protons the probabilities are somewhat lower since then $\beta < 1$.
Secondaries (*b*)	1·5	0·21	0·133	1, 700±250†	—	?		0·52	+	78	
(*c*)	1·6	0·54	2·07	600±120	—	Electron	$\sim$1	1·24	−	96	
(*d*)	1·5	0·14	0·56	230±30	—	Electron	$\sim$1	0·86	+	70	
11. Proton–proton collision.... (*a*)	0·5	1·2	0·44	150±15	60±3*	Proton	0·5	1·46	+	59	
(*b*)	0·9	2·2	1·00	90±9	90±4	Proton	0·42	1·20	+	67	
(*c*)	0·9	4·5	2·46	40±4	120±4	Proton	0·31	2·03	+	61	
12. Meson approaching end	2·1	10·9	13·7	20±2	Increasing	Meson	0·5	1·15	−	88	
13. Track leaving star	1·2	3·4	3·76	50±5	48±3	Meson	0·7	1·61	−	81	
14. Track leaving star	0·6	3·3	2·23	50±8	42±2	Meson	0·7	1·61	+	73	
15. Track leaving star	0·76	0·5	0·2	400±20	47±2	Proton	0·7	0·75	+	70	
16. Primary of star ..	0·55	0·41	0·1	500±100	33±2	Meson	$\sim$1	0·44	−	62	
17. Primary of star ..	1·2	0·77	3·12	300±25	65±3	Proton	0·7	5·8	+	97	

(2) *The Determination of the Sign of some Particles in Cosmic Ray Events.*

In the plates exposed at high altitude measurements have been made on several tracks of which in some cases the sign and the direction of motion can be deduced from the nature of the event. In others, such as electron pairs, conservation of charge must be maintained. These two groups can eventually be used as a verification of the efficiency of the method when sufficient statistics are available. When either or both the sign and the direction of motion of the particle are unknown a priori, the measurement of the sense of the deflection can be used in the analysis of the event.

Fig. 12

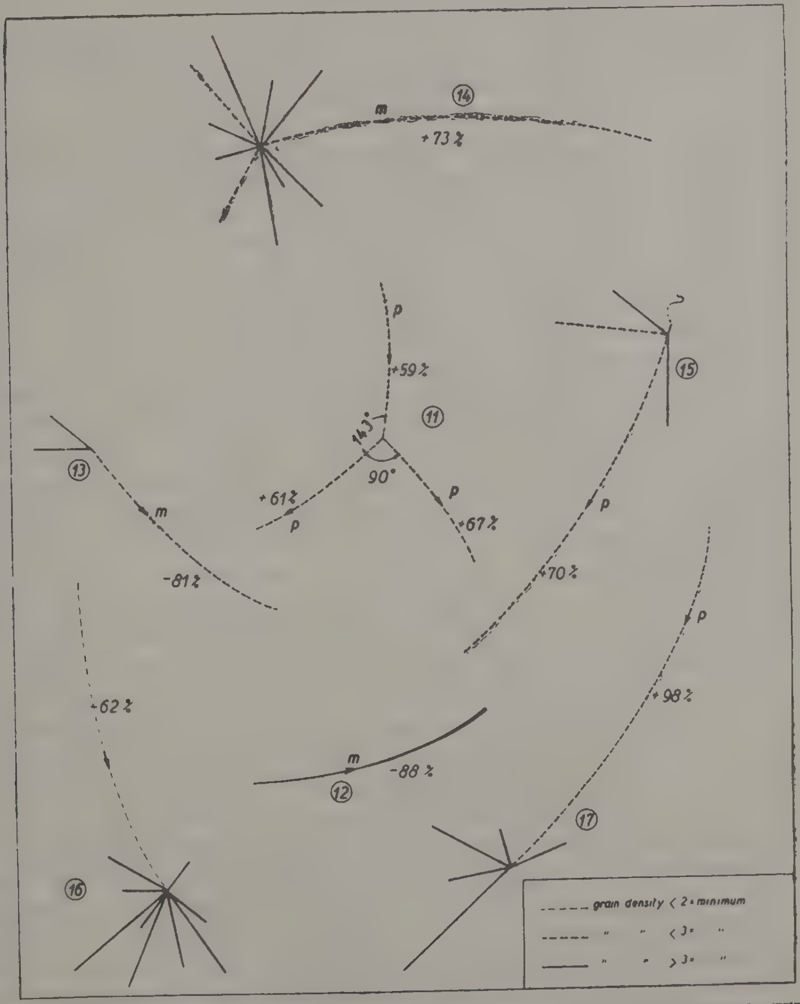


Diagramatic representation of the results presented in detail in Table III.

These tracks are listed in Table III. and sketched diagrammatically in figs. 12 and 13.

Track 1, identified as a proton, links two stars, one of which is evidently of higher energy and is therefore assumed to be the primary. This gives the direction of motion of the particle. Its sign is found to be positive with a probability of 95 per cent. This event is shown in the mosaic of Pl. XXXIV.

Fig. 13



Diagramatic representation of the results presented in detail in Table III.

Track 2, also a proton, gives rise to a knock-on electron, from which we can deduce the direction of motion. There is a probability of 93 per cent that the sign of this proton should be positive.

A branch of a star, track 3, is identified as a proton, its orientation and energy indicate that it comes from the star. It is found to be positive with a probability of 95 per cent.

Another proton, track 4, is the only high energy branch of a 4-branch star and the only one near the vertical. If it is the primary of the star it is positive, with a probability of 69 per cent.

Event 5 is the electron of a π–μ–e decay, it is found to be positive with a probability of 62 per cent. It was found in a corner of the plate, where the field was only 25,000 gauss.

Next there is a group of γ-ray pairs, of which the particles have energies ranging from 29 to 220 MeV. In each case one track of the pair is found to be positive, and the other negative, as should be expected.

Event 10 is a " giraffe " (G. Occhialini 1949). Since this is the direct production of a positron–electron pair by a charged primary, we expect to find the charge of the primary conserved among the three secondary particles. In fact, we find the primary to be positive with 83 per cent probability ; the secondaries to be two positive, one with 78 per cent probability, the other with 70 per cent probability; and the third negative with 96 per cent probability. There is also a rough energy balance, within the errors of measurement.

Event 11 seems to be a proton–proton collision at high energy. If this is so, the track coming from the vertical, a proton of 150 MeV. energy, is positive with a probability of 59 per cent; the continuation and recoil tracks are positive, one with 67 per cent and the other with 61 per cent probability. The angle of collision is 90°, the energy of the primary is divided between the two secondary tracks in a ratio in agreement with the angles of deflection.

Track 12 is of a meson which, by the increase of its grain density and scattering, can be seen to be approaching the end of its range. Its average energy is 20 MeV., its sign negative with a probability of 88 per cent.

The orientation of track no. 13, a meson by its grain density and scattering, is such as to make it likely that it comes out of the star; if this is so it is negative, the probability of the deflection being 81 per cent.

Track 14, also a meson, is moving nearly horizontally, but its energy is so low that it seems unlikely to be the primary of so large a star. Taking it to be a secondary, it is positive with 73 per cent probability.

In event 15 it is difficult to tell whether the track is a primary or secondary from either its orientation or energy. Since, however, it can be identified as a proton, and the deflection is such as to make it 70 per cent positive if it leaves the star, we accept it as a secondary.

Track 16 seems, by orientation and energy, to be the primary of a star of 10 low energy branches; it is a meson and has 62 per cent probability of being negative.

Event 17 is a star of many low energy branches, with one higher energy track coming from the vertical. Its energy and grain density indicate that it is a proton. It is positive with a probability of 98 per cent.

The average value of γ deduced from these measurements is $1{\cdot}56 \pm {\cdot}33$.

§ VI. DISCUSSION.

The results obtained seem to show that the use of a magnetic field to deflect tracks in the photographic plate can already contribute substantially to the information available about the cosmic ray events recorded. We discuss here the improvements which might be made to the technique in order to exploit to the full its possibilities.

Favourable conditions for the determination of the sign and momentum are a large value of the product $\gamma_0\beta\sqrt{t}$ and a low level of distortion in the emulsion. γ_0 can be increased by reducing the scattering constant of the emulsion. Already, as announced by Dr. Berriman of Kodak Ltd., and by Mr. Dodd and Mr. Waller of Ilford Ltd., at the Bristol Photographic Conference (May 1950), it has been possible to produce diluted emulsions containing a smaller proportion of silver bromide. An increase of 16 per cent in the theoretical value of γ_0 is obtained without loss of sensitivity.

While such plates present new problems in processing, the advantage of increasing γ_0 in this way is such as to warrant efforts to overcome the difficulties and to find emulsions with still smaller proportions of heavy elements.

The constant γ_0 may also be increased by increasing the field H, the mean track length by increasing the emulsion thickness, d. It can be shown that the number of tracks per unit area of plate for which $\gamma_0\sqrt{t}$ has a certain minimum value is proportional to H^2 and to d^2. In this sense, therefore, an increase in the thickness is as valuable as an increase in field. Moreover, while it is often of interest to examine single measurable tracks which emerge from a star, it is generally desirable to measure several tracks from the same event. If the directions of the tracks are independent, it can be shown that the number of events per unit area from which n tracks are all measurable is proportional to H^{2n} and to d^{n+1}.

For these reasons, throughout the experiments we have advanced cautiously towards thicker emulsions, and we are now scanning a batch of plates 1200 microns in thickness which show very little distortion. The table includes events taken from one of these plates. Further advances towards thicker emulsions are still possible.

The main difficulty in increasing the field is the high cost. Fields of 100,000 gauss or more are, in principle, quite possible, and the design of such a magnet is being investigated. Thick emulsions are more economical since they reduce the proportion of the field volume which is occupied by glass.

At the very highest momenta even the little information available about momentum from the magnetic deflection may be of value. The relative error of the momentum estimate by the magnetic deflection remains independent of momentum until the distortion becomes important. The distortion over a 1 cm. track in a 600μ plate has been shown in many cases to be as low as $1'$, and a correction may be applied even for this. We regard the upper limit of detectable momentum with this length of track as fixed by this angle, corresponding, in a field of 34,000 gauss, to a momentum of $3{\cdot}5\times10^{10}$ eV./c. This maximum is roughly proportional to track length and to the field. It is conceivable that measurements in the region above $10''$ eV./c may be possible if fields of the order of 100,000 gauss or more become available.

Fields of this magnitude will also make possible a determination of the mass of particles of velocity appreciably less than that of light, by comparison between scattering and magnetic deflection, which yields a value of β. Determination of mass by this method has recently been described for very low energy light particles in the cloud chamber (Groetzinger *et al.* 1950).

ACKNOWLEDGMENTS.

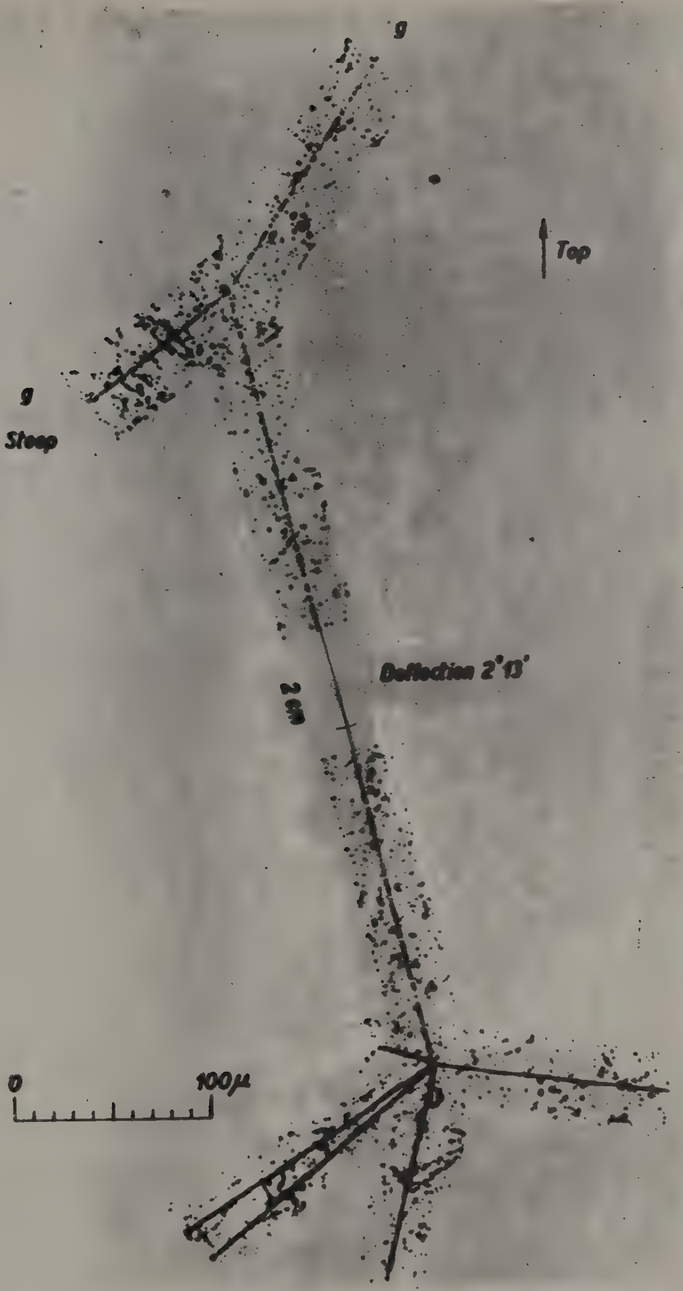
This paper describes part of the work being carried out in collaboration between the Photographic Emulsion groups of the Brussels Centre de Physique Nucléaire and the Manchester University Department of Physics on the magnetic deflection of cosmic-ray particles in the photographic emulsion. We wish to thank Professor P. M. S. Blackett for his constant interest in the work, and for the use of the magnet at the Pic-du-Midi. We are deeply indebted to Professors M. G. E. Cosyns and G. P. S. Occhialini and to Dr. G. D. Rochester for continual help and guidance during the work. Mr. J. E. Moyal has at all times been most helpful in discussion of theoretical problems. We wish to thank him also for communicating to us many of his results before publication. We wish to express our gratitude to our colleagues, Dr. G. Cortini, Dr. N. Page and Mr. G. Vanderhaeghe for their valuable assistance in the processing of the plates and in part of the measurements.

We wish to thank Professor Mailvaganam of the University of Colombo, for the use of their magnet and the research staff of Metropolitan Vickers Ltd., for operating the magnet during the exposure of the Kodak plates; we are especially grateful to Mr. P. P. Sterling and Mr. A. M. Armour. Mr. J. A. Newth kindly helped us in the design of the pole pieces. Dr. J. G. Wilson also gave us helpful advice in the early stages of the work.

The photographic plates were supplied by Dr. Berriman, of Messrs. Kodak Ltd., and Mr. Waller, of Messrs. Ilford Ltd., to whom we express our thanks. The problems connected with the adhesion and elimination of stress in the production of large plates of great thickness were not easy to overcome. We wish to express our gratitude to Mr. Waller and the staff of Ilford Ltd., who made possible this extension of the technique.

Through the friendly cooperation of Professor J. Svedberg and his staff a parallel sea-level exposure was made on the big magnet at the University of Upsala. The plates from this exposure were, unfortunately, subsequently found to have severe distortion and were therefore not used.

The exposures with the Ilford plates were made on the Blackett electromagnet at the Pic du Midi, France. We are deeply indebted to Dr. Rösch and his staff at the Pic du Midi for their help in the transport


g
Top
g
Steep
2 cm
Deflection 2°13'
0
100μ

Mosaic of event 1 of Table III. The track linking the two stars is shown to be positive with a probability 95 per cent, the particle being a proton. The direction of motion of the particle is determined from considerations of the energy of the two stars, and from orientation.

To face page 1056

and erection of the magnet and for their unfailing kindness and courtesy. We wish also to express our thanks to Dr. Butler and to Mr. K. H. Barker and the late Mr. C. O. Green for the transportation and erection of the magnet at the Pic du Midi.

REFERENCES.

BLACKETT, 1936, *Proc. Roy. Soc.*, Ser. A, **154,** 564.
BRADDICK, 1950, *J. Sci. I.* To be published.
CAMERINI, COOR, DAVIES, FOWLER, LOCK, MUIRHEAD and TOBIN, 1949, *Phil. Mag.*, **40,** 1073.
COSYNS and VADERHAEGHE, 1950, Note No. 15, *Bull. Cen. de Phys. Nuc., de l'Université Libre de Bruxelles ;* Bristol Congress on the Photographic process.
COSYNS, 1949, Paper read at Como Congress.
DILWORTH, OCCHIALINI and PAYNE, 1948, *Nature*, **162,** 102.
DILWORTH, OCCHIALINI and VERMAESEN, 1950, Note no. 13, *Bull. du Cen. de Phys. Nuc. de l'Université Libre de Bruxelles.*
FOWLER, 1950, *Phil Mag.*, **41,** 169.
GOLDSCHMIDT-CLERMONT, KING, MUIRHEAD and RITSON, 1948, *Proc. Phys. Soc.*, **61,** 183.
GOLDSCHMIDT-CLERMONT, 1950, Note No. 14, *Bull. du Cen. de Phys. Nuc. de l'Université Libre de Bruxelles ;* Nuovo Cimento, **7,** 331.
GROETZINGER, BERGER and LIBE, *Phys. Rev.*, **77,** 584.
MOYAL, 1950, *Phil. Mag.*, **41,** 1058.
OCCHIALINI, 1949, Paper read at Como Congress.
SCOTT, 1949, *Phys. Rev.*, **76,** 213.

XCV. *The Momentum and Sign of Fast Cosmic Ray Particles.*

By J. E. Moyal,
Department of Mathematics, The University, Manchester *.

Summary.

In order to study the statistical estimation of momentum and sign of fast cosmic ray particles from their magnetically deflected and multiply scattered tracks in the photographic emulsion, the probability distribution of the observed angular deviations along the track is derived, taking into account errors of observations. With the help of this distribution, approximate expressions are obtained for the maximum likelihood estimates of the magnetic deviation and the mean-square scattering deviation. The standard errors of these estimates are found, and their efficiency is discussed.

After showing how the observations of magnetic and scattering deviations in a large number of tracks can be combined to verify theoretical mean-square scattering formulæ, the estimation of the momentum p of particles from these deviations is discussed. An expression is derived for the value of the optimum cell-length, at which the error of the scattering determination of p is minimized. The relative errors of the magnetic and scattering determinations of p are compared at this optimum cell-length : it is found that there is a critical value of p, depending on the field, below which the scattering determination is more accurate.

An expression is given for the probability that, given the signed total deviation of a track in the magnetic field, the particle should be positively or negatively charged. A method is given of estimating the ratio of positively to negatively charged particles in terms of the observed signed deviations in a number of tracks.

§1. Introduction.

Fast charged particles passing through the emulsion of photographic plates suffer multiple scattering due to elastic collisions with the nuclei in the emulsion. If a uniform magnetic field is applied normal to the plate, a constant angular deviation per unit track length is superimposed on this random scattering. Estimates of the momentum of a particle may be obtained either from the mean-square angle of scattering (*cf.* Goldschmidt-Clermont *et al.* 1948), or from the mean curvature due to magnetic forces ; estimates of its *sign*, however, are only possible from the magnetic deviation, provided, of course, we know the direction of travel of the particle when it enters the field.

As the expected mean values of magnetic deviation due to fields of 20,000 to 250,000 gauss is of the same order of magnitude as the scattering angle and the spurious apparent deviations due to finite grain size in the emulsion, errors of observation, etc., the statistical problem of estimation from photographic plate data becomes important. The purpose of the present paper is to develop the theory of multiple scattering with a superimposed magnetic deviation in a form suitable for the study of estimates of the signs and momenta of particles from data contained in their tracks, and to evaluate the relative merits of such estimates and the errors affecting them.

---

* Communicated by Professor P. M. S. Blackett, F.R.S.

§ 2. MOTION OF MULTIPLY SCATTERED PARTICLES IN A UNIFORM MAGNETIC FIELD.

In this section the theory of multiple scattering is expressed in a form convenient for the solution of statistical estimation problems. We shall be concerned with the " true " motion only: the effects of errors of observation will be considered in the next section. We limit ourselves to particles fast enough for the loss of energy to be negligible in any given track, so that the velocity $\boldsymbol{\beta}$ has very approximately a constant magnitude β; the letter t will therefore be used to denote both time divided by β and distance along the track. We shall use the units common in cosmic ray work: velocity will be measured in multiples of the velocity of light c, charge in multiples of the electron charge, energy in electron-volts, momentum in eV./c., length in cm.; magnetic fields, however, will be in gauss (a unit consistent with the previous ones would be V/cm.c.).

The quantum mechanical treatment of scattering by nuclei of relativistic charged particles, such that nuclear interactions and the recoil of nuclei can be neglected, gives the following results (Williams 1939; *cf.* also Molière 1947):

(1) The cross-section for single scattering is very approximately the Coulomb cross-section modified to account for screening and finite size of the nucleus. Roughly speaking, the effect of screening is to introduce a lower limit to the scattering angle, while the finite size of the nucleus introduces an upper limit.

(2) Successive collisions may be treated as independent events; in a dense scatterer, the distribution of the total angle through which a particle scatters in an interval δt long enough for a large number N of collisions to occur is, therefore, the resultant of the composition of N independent modified Coulomb scattering distributions. The mean-square of the angular deviation $\delta\phi$ projected on a plane is then shown to be proportional to δt:

$$\overline{\delta\phi^2}=s^2\delta t=\left(\frac{\nu}{p}\right)^2\delta t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

(the bar over a symbol denotes its expectation value); the mean-square deviation per unit track s^2 is shown to be inversely proportional to the square of the momentum p, the coefficient ν *depending only on the properties of the scatterer*. For either Ilford G5 or Kodak NT4 electron sensitive plates the calculation of ν on the basis of Williams' approximate formulæ gives $\nu=8{\cdot}48\times10^6$ (with $\delta\phi$ in radians). As δt increases, the distribution of $\delta\phi$ tends rapidly to the Gaussian in its central part, but maintains a longer "tail" to begin with, due to an appreciable probability of occasional close collisions with large scattering (*cf.* Williams, *loc. cit.*; see also Goudsmit and Saunderson (1940), Molière (1948), Snyder and Scott (1949), and Lewis (1950) for a more complete treatment; the application to photographic plates is considered by Goldschmidt-Clermont (1950). In particular, for the order of intervals and range of momenta of interest in this paper, one may assume (as we shall do henceforth) a Gaussian distribution for $\delta\phi$, provided one excludes these occasional large-angle single scatterings from the observations, *e.g.* by "cutting" the observed distribution at its mean $\pm$ 3 times its standard deviation*.

What we shall require for later work is the joint distribution of angular deviations and coordinates of particles for tracks selected to lie approximately in a plane (the emulsion plane). This may be obtained by solving the appropriate "diffusion" equation which this distribution function obeys (*cf.* Fermi, as quoted by Rossi and Greisen (1941)); it is convenient, however, when dealing with the estimation problem, to treat directly the equations of motion for the angles and coordinates themselves †. Let then θ, ϕ be the polar angles defining the direction of the particle velocity $\boldsymbol{\beta}$, with the emulsion in the equatorial or (x, y) plane. The relativistic equations of motion for a multiply scattered particle in the presence of a uniform magnetic field H normal to the emulsion plane lead to the following expressions for the increments $\delta\theta$, $\delta\phi$ of the polar angles in a small interval δt:

$$\delta\theta=\delta\Theta;\quad \sin\theta\,\delta\phi=\omega\sin\theta+\delta\Phi, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where

$$\omega=\frac{300\,e\mathrm{H}}{p} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

is the magnetic deviation per unit track (e being the charge, p the momentum of the particle), and $\delta\Theta$, $\delta\Phi$ are the projections of the scattering angular deviations in δt on two orthogonal planes containing $\boldsymbol{\beta}$, the first containing the z-axis as well. Owing to the symmetry of the scattering distribution about the velocity at the beginning of the interval, $\delta\Theta$ and $\delta\Phi$ must be independent with zero means and equal mean squares, so that

$$\overline{\delta\Theta^2}=\overline{\delta\Phi^2}=s^2\delta t. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The formal integration of (2) from time 0 to t gives

$$\phi(t)=\phi_0+\omega t+\int_0^t\frac{d\Phi(\tau)}{\sin\theta(\tau)},$$

---

* See Dilworth *et. al.* (1950) for further discussion of effect of cut-off.

† These correspond to the Langevin equations for the Brownian motion, while the diffusion equation corresponds to the Fokker-Planck equation (see, *e. g.* Moyal 1949).

where ϕ_0 is the azimuth at $t=0$. Since we are selecting tracks lying approximately in the (x, y) plane, $\theta=\left(\frac{\pi}{2}\pm\epsilon\right)$, where $\epsilon<5^0$ say. Neglecting, then, terms of order ϵ^2, we have

$$\phi(t)=\phi_0+\omega t+\int_0^t d\Phi(\tau)=\phi_0+\omega t+\Phi(t). \quad . \quad . \quad . \quad . \qquad (5)$$

The velocity components along the x- and y-axes respectively are

$$\left.\begin{aligned}\beta_x(t)/\beta&=\cos\phi=\cos(\phi_0+\omega t+\Phi)\simeq\cos\phi_0-(\omega t+\Phi)\sin\phi_0,\\ \beta_y(t)/\beta&=\sin\phi=\sin(\phi_0+\omega t+\Phi)\simeq\sin\phi_0+(\omega t+\Phi)\cos\phi_0,\end{aligned}\right\} \quad . \qquad (6)$$

where the last approximate expressions in the right-hand sides of equations (6) are valid as long as both magnetic and scattering deviations remain small (both $s\sqrt{t}$ and $\omega t<5^0$ say). Integrating, we obtain for the coordinates of the particle

$$\left.\begin{aligned}x(t)&=x_0+t\cos\phi_0-\tfrac{1}{2}\omega t^2\sin\phi_0-\sin\phi_0\int_0^t(t-\tau)d\Phi(\tau),\\ y(t)&=y_0+t\sin\phi_0+\tfrac{1}{2}\omega t^2\cos\phi_0+\cos\phi_0\int_0^t(t-\tau)d\Phi(\tau),\end{aligned}\right\} \qquad (7)$$

where x_0, y_0 are the coordinates at $t=0$.

Equations (5) and (7) are valid in the small angle approximation, whatever the distribution of $\delta\Phi$. If the latter is Gaussian, then it can be shown (see *e. g.* Moyal 1949) that the joint distribution of ϕ, x and y will be of the *multivariate Gaussian* type, whose general expression for n variables $z_1, \ldots, z_n$ (which we shall require again later) may be written succinctly as follows (*cf.* Kendall 1946 or Cramèr 1946): let the vector $\mathbf{z}$ stand for the n variables, $\mathbf{m}$ for their means, $\mathbf{\Lambda}$ for their *covariance matrix*, whose elements are

$$\lambda_{ij}=\overline{(z_i-m_i)(z_j-m_j)},$$

$\mathbf{\Lambda}^{-1}$ for the inverse of $\mathbf{\Lambda}$ and Λ_n for its determinant; the probability that $\mathbf{z}$ should lie in the n-dimensional volume element $dz_1 \ldots dz_n$ is then

$$\mathrm{P}(\mathbf{z})dz_1\ldots dz_n=(2\pi)^{-\frac{1}{2}n}\Lambda_n^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(\mathbf{z}'-\mathbf{m}')\mathbf{\Lambda}^{-1}(\mathbf{z}-\mathbf{m})\}dz_1\ldots dz_n. \qquad (8)$$

$\mathrm{P}(\mathbf{z})$ is completely determined by the means and covariances of the variables. In the present case the means are obviously (since $\delta\overline{\Phi}=0$)

$$\left.\begin{aligned}m_\varphi&=\overline{\phi}=\omega t,\\ m_x&=\overline{x}=x_0+t\cos\phi_0-\tfrac{1}{2}\omega t^2\sin\phi_0,\\ m_y&=\overline{y}=y_0+t\sin\phi_0+\tfrac{1}{2}\omega t^2\cos\phi_0.\end{aligned}\right\} \quad . \quad . \quad . \qquad (9)$$

The calculation of the variances and covariances is simplified by the fact that the scattering deviations over non-overlapping intervals are independent, so that formally

$$\overline{d\Phi(\tau)d\Phi(\tau')}=s^2\delta(\tau-\tau')d\tau d\tau',$$

where $\delta(\tau)$ is Dirac's delta-function. Hence, *e. g.*

$$\int_0^t\int_0^t (t-\tau)(t-\tau')\overline{d\Phi(\tau)d\Phi(\tau')}=s^2\int_0^t (t-\tau)^2 d\tau=\tfrac{1}{3}s^2t^3,$$

and similarly for other terms involving the expectation values of products of such integrals. Finally, we find

$$\left.\begin{aligned}
\lambda_{\varphi\varphi}&=\overline{\phi'^2}=s^2t,\\
\lambda_{xx}&=\overline{x'^2}=\tfrac{1}{3}s^2t^3\sin^2\phi_0,\\
\lambda_{yy}&=\overline{y'^2}=\tfrac{1}{3}s^2t^3\cos^2\phi_0,\\
\lambda_{\varphi x}&=\overline{\phi'x'}=\tfrac{1}{2}s^2t^2\sin\phi_0,\\
\lambda_{\varphi y}&=\overline{\phi'y'}=\tfrac{1}{2}s^2t^2\cos\phi_0,\\
\lambda_{xy}&=\overline{x'y'}=\tfrac{1}{3}s^2t^3\sin\phi_0\cos\phi_0,
\end{aligned}\right\}\quad . \quad . \quad . \quad . \quad . \quad (10)$$

where we have written $\phi'=\phi-\overline{\phi}$, etc. With the origin of coordinates at x_0, y_0, and the x-axis along the initial direction of the particle ($\phi_0=0$) so that in the small angle approximation $\lambda_{xx}=0$ and $x=t$, the distribution reduces to that of ϕ and y only :

$$\mathrm{P}(\phi, y, t)$$

$$=\frac{\sqrt{3}}{\pi s^2t^2}\exp\left\{-\frac{2}{s^2t^3}\left[t^2(\phi-\omega t)^2-3t(\phi-\omega t)(y-\tfrac{1}{2}\omega t^2)+3(y-\tfrac{1}{2}\omega t^2)^2\right]\right\}.$$

$$. \quad . \quad . \quad (11)$$

Fermi's solution of the diffusion equation (Rossi and Greisen, *loc. cit.*) is obtained from (11) by making $\omega=0$.

§ 3. THE DISTRIBUTION OF OBSERVED ANGULAR DEVIATIONS.

The fundamental quantities that we have to estimate from the data in the tracks are : (a) the magnetic deflection, and (b) the mean-square scattering deviation, *i.e.* in the notation of § 2, the parameters ω and s^2. Two experimental methods have been used for this purpose : in the first, which we may call the *angular difference method* (Goldschmidt-Clermont *et. al.* 1948), the eyepiece goniometer of a microscope is rotated until a straight line graticule is in line with the track, and its angle with a fixed reference line is noted. The operation is repeated along the whole length of the track, advancing each time the plate (which is mounted on a stage movement with micrometer adjustment) by a fixed distance, called the *cell-length*. The estimations are made from the angular differences between successive cells. In the second method, which may be called the *coordinate method* (Fowler 1950), the track is aligned approximately parallel to one of the directions of motion of the stage, say the x-axis and the y-coordinate at which the track intersects a graticule in the eyepiece is determined by means of an eyepiece scale. These measurements are repeated at every cell-length along the whole track, as in the first method. Clearly, this method could be extended to measure the coordinates of every blackened grain : this would yield all the information contained in the track. Cell-lengths would then cease to be constant, but would be equal to the distances between successive grains : apart from this last added complication, the estimation problem would remain unchanged. Methods based on the same principles may, of course, be used in conjunction with a projection microscope.

The first step in formulating the problem of statistical estimation is to derive the probability distribution of the actual observation, taking into account not only scattering, but also errors of various kinds. This requires a closer analysis of the actual methods of observation : we consider first the angular difference method. When aligning the eyepiece graticule along the track, the observer has to use his judgment to determine the best straight line through the grains in each cell (for a particle at minimum ionization, the grain density is of the order of 33 in a 100-micron cell). The line taken may be expected to lie somewhere between the tangent at midpoint and the chord in each cell (smoothing effect). There will, of course, be a random error due (*a*) to observer and instrumental error, and (*b*) to finite grain size and to the asymmetry of grain distribution along the trajectory : the centres of developed grains may be slightly displaced from the actual trajectory. In addition, there is the possibility of systematic errors due to distortion of the emulsion. Special precautions are taken to eliminate this last effect : we shall therefore ignore it in our analysis. Here we shall assume that at each measurement the observer aligns the graticule along a line which divides the angle between chord and tangent in a fixed ratio q, with a total random error ϵ, due to various causes enumerated above. (We shall call ϵ the *spurious deviation* and q the *smoothing factor*.) We also assume that ϵ has a Gaussian distribution, with zero mean and the same standard deviation s_0 in every cell. These assumptions seem to be well-founded on the basis of the careful study of sources of error made by Goldschmidt-Clermont *et. al* (1948 and 1950).

We proceed now to derive the required probability distribution of the observed angular deviations. In a track of total length t, let h be the length of each cell, $n+1$ the number of cells, so that $(n+1)h=t$: we shall assume throughout in what follows that n is large, *e. g.* >25. We take the track to be initially aligned along the x-axis, and define all angles relative to this axis. Let then ψ_r be the *observed* angle so defined of the rth cell, ϕ_r the angle of the tangent to the *true* trajectory at the midpoint of this cell, α_r the angle of the corresponding chord :

$$\alpha_r \simeq (y_{r+1}-y_r)/h,$$

where y_r, y_{r+1} are the end coordinates of the cell ; ϵ_r is the error of observation in the setting of ψ_r. According to the foregoing assumptions

$$\left.\begin{aligned} \psi_r &= \alpha_r + q(\phi_r-\alpha_r)+\epsilon_r, \\ \delta\psi_r &= \psi_{r+1}-\psi_r = q\delta\phi_r+(1-q)\delta\alpha_r+\epsilon_{r+1}-\epsilon_r. \end{aligned}\right\} \quad . \quad . \quad . \quad (1)$$

Each angular deviation $\delta\psi_r$ is thus a sum of Gaussian variables, and is therefore itself Gaussian. Hence the joint distribution of the observations (known as the *likelihood function*) will be the Gaussian multivariate distribution, as defined by equation (8) of § 2, the vector $\mathbf{z}$ standing now for the n deviations $\delta\psi_1, \ldots, \delta\psi_n$; all we have to do in order to find the explicit expression of this distribution in terms of the parameters we wish to estimate, *i. e.* of ω, s and s_0, is to express the means, variances, and covariances of the deviations in these terms. This is easily achieved by the methods of § 2, with the following results : the means are

$$m_r = \overline{\delta\psi_r} = \omega h, \quad (r=1, 2, \ldots, n) ; \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

the covariances are (writing $\delta\psi_r' = \delta\psi_r - \overline{\delta\psi_r}$)

$$\left.\begin{aligned} \lambda_{r,r} &= \overline{\delta\psi_r'^2} &&= \tfrac{2}{3}(1-\tfrac{1}{2}q+q^2)s^2h+2s_0^2, \\ \lambda_{r,r+1} &= \overline{\delta\psi_r'^2\delta\psi_{r+1}'} &&= \tfrac{1}{6}(1+q-2q^2)s^2h-s_0^2, \\ \lambda_{r,r+k} &= \overline{\delta\psi_r'\delta\psi_{r+k}'} &&= 0, \quad (k>1 ; r=1, 2, \ldots, n). \end{aligned}\right\} \quad . \quad . \quad (3)$$

Note that the sum

$$\lambda_{r,r} + 2\lambda_{r,r+1} = s^2h \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

is independent of the " smoothing factor " q. If $q=1$, then

$$\lambda_{r,r} = s^2h+2s_0^2 \quad \text{and} \quad \lambda_{r,r+1} = -s_0^2, \quad (r=1, 2, \ldots, n). \quad . \quad . \quad . \quad (5)$$

In the coordinate method we may expect the random error affecting each coordinate measurement to arise, as in the angular difference method, (a) from observer and instrumental errors, and (b) from finite grain size and asymmetry of grain distribution about the track ; there will, however, be no equivalent to the " smoothing effect". It is reasonable, therefore, to regard the rth measured coordinate as the sum of the " true " coordinate y_r and of a total random error ϵ_r, and to assume that the ϵ_r all have the same Gaussian distribution with zero mean and standard deviation s_0. Since this method can be applied to the measurement of coordinates of every developed grain, we shall develop the theory *ab initio* for unequal cell-lengths.

Let h_r be the rth cell-length and $n+1$ the number of cells in a track of total length t, so that $\sum_1^{n+1} h_r = t$. It is convenient to form the estimates in terms of second-differences of the *observed* coordinates, corresponding to the angular differences of the previous method : we write, therefore,

$$\delta\psi_r = \frac{y_{r+2}-y_{r+1}}{h_{r+1}} - \frac{y_{r+1}-y_r}{h_r} + \frac{\epsilon_{r+2}-\epsilon_{r+1}}{h_{r+1}} - \frac{\epsilon_{r+1}-\epsilon_r}{h_r}. \quad . \quad . \quad . \quad (6)$$

Once again these deviations are sums of Gaussian variables, and, therefore, Gaussian themselves : their joint distribution is hence again of the form (8), *cf.* § 2. Calculation of the means, variances and covariances gives the following results :

$$\left.\begin{aligned} m_r&=\overline{\delta\psi_r}=\tfrac{1}{2}\omega(h_{r+1}+h_r),\\ \lambda_{rr}&=\overline{\delta\psi_r'^2}=\tfrac{1}{3}s^2(h_{r+1}+h_r)+2s_0^2\left(\frac{1}{h_{r+1}^2}+\frac{1}{h_{r+1}h_r}+\frac{1}{h_r^2}\right),\\ \lambda_{r,r+1}&=\overline{\delta\psi_r'\delta\psi_{r+1}'}=\tfrac{1}{6}s^2h_{r+1}-s_0^2\left(\frac{1}{h_{r+2}h_{r+1}}+\frac{1}{h_{r+1}^2}+\frac{1}{h_{r+1}h_r}\right),\\ \lambda_{r,r+2}&=\overline{\delta\psi_r'\delta\psi_{r+2}'}=s_0^2\left(\frac{1}{h_{r+2}h_{r+1}}\right),\\ \lambda_{r,r+k}&=\overline{\delta\psi_r'\delta\psi_{r+k}'}=0, \qquad (k>2\ ;\ r=1,2,\ldots,n). \end{aligned}\right\} \quad . \quad (7)$$

Note that in consequence

$$\lambda_{r,r}+\lambda_{r,r-1}+\lambda_{r,r+1}+\lambda_{r,r-2}+\lambda_{r,r+2}=\tfrac{1}{2}s^2(h_r+h_{r+1}). \quad . \quad . \quad (8)$$

§ 4. MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS.

What we must do now is to establish suitable statistics giving s and ω in terms of the observed deviations ; it will be necessary, incidentally, to estimate s_0 as well. We use for this purpose the *maximum likelihood method* (*cf.* Kendall 1946, Cramer 1946), *i. e.* we maximize the probability of the whole set of observations, obtained in the last section, with respect to the parameters, ω, s and s_0 considered as unknowns. In this section we shall give a formal solution of the problem, which covers the angular difference and the coordinate methods.

Let **h** stand for the cell lengths $h_1, \ldots, h_n$; the means vector is then

$$\mathbf{m}=\omega\mathbf{h}.$$

The logarithm of the likelihood function is (*cf.* equation (8) of § 2)

$$\mathrm{L}=\log \mathrm{P}=-\frac{n}{2}\log 2\pi-\tfrac{1}{2}\log \Lambda_n-\tfrac{1}{2}(\mathbf{z}'-\omega\mathbf{h}')\,\mathbf{\Lambda}^{-1}(\mathbf{z}-\omega\mathbf{h}). \quad . \quad . \quad (1)$$

The covariance matrix $\mathbf{\Lambda}$ is a function of s and s_0 only (*cf.* equations (3) and (7) of § 3). It is convenient to express it in the form

$$\mathbf{\Lambda}(s, s_0)=\eta\mathbf{A}(\chi), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where η is a linear combination of s^2 and s_0^2, χ some function of s^2 and s_0^2, the elements of the matrix **A** are functions of χ only, and to maximize the likelihood function

$$\mathrm{L}=-\frac{n}{2}\log(2\pi\eta)-\tfrac{1}{2}\log \mathrm{A}_n-\frac{1}{2\eta}(\mathbf{z}'-\omega\mathbf{h}')\mathbf{A}^{-1}(\mathbf{z}-\omega\mathbf{h}) \quad . \quad . \quad (3)$$

(A_n being the determinant of $\mathbf{A}$) in terms of these new variables. Solving $\partial L/\partial\omega=0$, one obtains for the estimate $\hat{\omega}$ of the magnetic deviations per unit track

$$\omega=\frac{\mathbf{h}'\mathbf{A}^{-1}\mathbf{z}}{\mathbf{h}'\mathbf{A}^{-1}\mathbf{h}}. \qquad (4)$$

Similarly, the estimate $\hat{\eta}$ is obtained by solving $\partial L/\partial\eta=0$:

$$n\hat{\eta}=\boldsymbol{\xi}'\mathbf{A}^{-1}\boldsymbol{\xi}, \qquad (5)$$

where $\boldsymbol{\xi}=\mathbf{z}-\omega\mathbf{h}$. In order to estimate χ we write

$$\frac{\partial L}{\partial\chi}=-\frac{1}{2}\frac{\partial\log A_n}{\partial\chi}-\frac{1}{2\eta}\frac{\partial}{\partial\chi}(\boldsymbol{\xi}'\mathbf{A}^{-1}\boldsymbol{\xi})=0,$$

or, substituting for η from (5),

$$\frac{\partial}{\partial\chi}\{A_n^{1/n}\boldsymbol{\xi}'\mathbf{A}^{-1}\boldsymbol{\xi}\}=0. \qquad (6)$$

Substituting for ω from (4), (6) becomes a transcendental equation in χ only. The substitution of its solution $\hat{\chi}$ in (4) and (5) gives the estimates $\hat{\omega}$ and $\hat{\eta}$; finally s^2 and s_0^2 are obtained from $\hat{\eta}$ and $\hat{\chi}$. Formally, this completes the solution of the problem. In practice (6) has to be solved by some approximation method : *e. g.* by choosing χ such that $|\chi|<1$, one can expand A_n and $\mathbf{A}^{-1}$ in powers of χ, and approximate to the left-hand side of (6) by a polynomial in χ (*cf.* the calculations in the next paragraph).

It has been shown in the case of independent observations (*cf.* Kendall, Cramér, *loc. cit.*) that the variances of the maximum likelihood estimates are asymptotically the smallest, in the sense that no other estimate can have a smaller variance in the limit when the number of observations $n\rightarrow\infty$. Though this proposition is not proved in general for correlated observations, we may yet expect it to hold in the present case, because the correlations extend only to first, or first and second, nearest neighbour. It is often found that these variances have comparatively simple expressions, even when the maximum likelihood estimates themselves are complicated. In such cases, their knowledge allows one to assess the *efficiency* of any simple (or approximate) estimate, defined formally as the limit when $n\rightarrow\infty$ of the inverse ratio of the variance of the latter to that of the corresponding maximum likelihood estimate.

The expression for the variances of the maximum likelihood estimates may be obtained as follows (*cf.* Kendall, *loc. cit.*) : let $\theta_1, \theta_2, \ldots$ be the parameters we wish to estimate. The covariance matrix $\mathbf{V}$ of the estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots$ is then the inverse of a matrix $\mathbf{V}^{-1}$, whose elements are

$$V_{ij}=-E\left\{\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right\}=-\int\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}P(\mathbf{z})d\mathbf{z} \qquad (7)$$

(E denoting expectation value). We are interested only in the variances of $\hat{\omega}$ and $\hat{s}^2$: we may start again from expression (3) for the likelihood function but with different parameters, substituting $\mathbf{B}$ for $\mathbf{A}$, s^2h for η, ξ for χ, and taking ξ such that the elements of $\mathbf{B}$ are linear in ξ. Write $\theta_1=\omega$, $\theta_2=s^2h$ and $\theta_3=\xi$. The following expressions are found for the elements of $\mathbf{V}^{-1}$:

$$\mathrm{V}_{11}=\frac{\mathbf{h}'\mathbf{B}^{-1}\mathbf{h}}{s^2h}\,;\ \mathrm{V}_{22}=\frac{n_4}{2s^4h^2}\,;\ \mathrm{V}_{33}=-\tfrac{1}{2}\frac{d^2\log \mathrm{B}_n}{d\xi^2}\,;$$

$$\mathrm{V}_{23}=\frac{1}{2s^2h}\frac{d\log \mathrm{B}_h}{d\xi}\,;\ \mathrm{V}_{12}=\mathrm{V}_{13}=0. \quad . \quad . \quad . \quad (8)$$

Finally, writing $\log \mathrm{B}_n=n\log b$, we obtain

$$\left.\begin{aligned}\text{var. }(\hat{\omega})&=\frac{1}{\mathrm{V}_{11}}=\frac{s^2h}{\mathbf{h}'\mathbf{B}^{-1}\mathbf{h}},\\ \text{var. }(\hat{s}^2h)&=\frac{\mathrm{V}_{33}}{\mathrm{V}_{22}\mathrm{V}_{33}-\mathrm{V}_{23}^2}=\frac{2s^4h^2}{n}\left\{1-\frac{(db/d\xi)^2}{b(d^2b/d\xi^2)}\right\}.\end{aligned}\right\} \quad . \quad . \quad (9)$$

§ 5. STATISTICAL ESTIMATES IN THE ANGULAR DIFFERENCE METHOD.

We proceed now to apply the results of the last paragraph to the estimation problem in the angular difference method; it is hoped to consider the coordinate method in a later paper. Write

$$\lambda_{r,r}=v, \qquad \lambda_{r,r+1}=-w, \qquad (r=1, 2, \ldots, n)$$

for the elements of the covariance matrix (equations (2) and (3) of § 3), and $\mathbf{\Lambda}=v\mathbf{A}$ (*i. e.*, we take $\eta=v$ in expression (2) of § 4), so that

$$\mathbf{A}=\begin{bmatrix} 1 & -\alpha & 0 & 0 \ldots\ldots 0 & 0 \\ -\alpha & 1 & -\alpha & 0 \ldots\ldots 0 & 0 \\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ 0 & 0 & 0 & 0 \qquad -\alpha & 1 \end{bmatrix} \quad . \quad . \quad . \quad (1)$$

where $\alpha=w/v$. It is easily seen that the determinant A_n of $\mathbf{A}$ satisfies the following difference equation (*cf.* Koopmans 1942):

$$\mathrm{A}_n=\mathrm{A}_{n-1}-\alpha^2\mathrm{A}_{n-2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

which is satisfied by a solution proportional to β^n, provided that

$$\beta^2-\beta+\alpha^2=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Finally, introducing as "initial" conditions the values of $\mathrm{A}_1=1$ and $\mathrm{A}_2=1-\alpha^2$, we find

$$\mathrm{A}_n=\frac{\beta_1^{n+1}-\beta_2^{n+1}}{\beta_1-\beta_2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where β_1, β_2 are the two solutions of (3) :

$$\beta_1=\tfrac{1}{2}(1+\sqrt{(1-4\alpha^2)})\,; \qquad \beta_2=\tfrac{1}{2}(1-\sqrt{(1-4\alpha^2)}). \quad . \quad . \quad . \quad (5)$$

It is easily shown that the elements A_{rs} of $\mathbf{A}^{-1}$ can be written

$$A_{rs}=\alpha^{s-r}\left(\frac{A_{r-s}A_{r-1}}{A_n}\right), \qquad (s\geqslant r). \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Write now $\lambda^2=\beta_2/\beta_1$; after a few simple transformations one finds

$$\left.\begin{aligned} \alpha&=\frac{\lambda}{1+\lambda^2}, \\ A_n&=(1-\lambda^2)^{-1}(1+\lambda^2)^{-n}(1-\lambda^{2(n+1)}), \\ A_{r,r+k}&=\frac{1+\lambda^2}{1-\lambda^2}\lambda^k(1-\lambda^{2r})(1-\lambda^{2(n-r-k+1)})(1-\lambda^{2(n+1)})^{-1}\,; \end{aligned}\right\} \quad . \quad (7)$$

A_n and $\mathbf{A}^{-1}$ have thus comparatively simple expressions as functions of λ only, and it will be convenient to maximize the likelihood function with respect to v and λ ; *i. e.* to take $\eta=v$ and $\chi=\lambda$ in (2) of § 4.

From their definition, one can see that $-\frac{1}{2}\leqslant w/v\leqslant\frac{1}{2}$, and hence $\alpha^2\leqslant\frac{1}{4}$, $\lambda^2\leqslant 1$. It is shown in the next section that optimum conditions are realized when $s_0/s\sim\frac{1}{2}$, and hence $\alpha\sim\frac{1}{6}$ and $\lambda^2\sim 0{\cdot}03$, so that near optimum $\lambda^k\ll 1$ for $k>2$: in what follows we shall therefore neglect $1/n$ and λ^k as compared to unity for $k>2p$, where $p\ll n/2$.

With these approximations, a straightforward calculation, using (7), gives

$$\sum_{r,s} A_{rs}=n\frac{1+\lambda^2}{(1-\lambda)^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

so that the estimate (4) of § 4 for the magnetic deviation becomes

$$\hat{\omega}t=n\frac{\sum\limits_{r,s} A_{rs}\delta\psi_s}{\sum\limits_{r,s} A_{rs}}=\psi-\frac{1}{1+\lambda}\sum_{k=1}^{2p}\lambda^k(\delta\psi_k+\delta\psi_{n-k+1}), \quad . \quad . \quad . \quad (9)$$

where $\psi=\sum\limits_{1}^{n}\delta\psi_k$ is simply the total observed deviation of the track ; the last term in the right-hand side of (9), representing what we may term an "end-effect" of the extreme observations, will generally be $\ll\psi$. Hence the first approximation to $\hat{\omega}$ will be just

$$\hat{\omega}_1 t=\psi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9\,a)$$

Similarly, (6) of § 4 gives for $\hat{v}$ (writing $\xi_k=\delta\psi_k-\omega h$)

$$n\hat{v}=\boldsymbol{\xi}'\mathbf{A}^{-1}\boldsymbol{\xi}=\frac{1+\lambda^2}{1-\lambda^2}\left\{\sum_{1}^{n}\xi_r^2+2\sum_{k=1}^{2p}\lambda^k\sum_{r=1}^{n-k}\xi_r\xi_{r+k}-\left(\sum_{k=1}^{p}\lambda^k\xi_k\right)^2\right.$$
$$\left.-\left(\sum_{k=1}^{p}\lambda^k\xi_{n-k+1}\right)^2\right\}, \quad . \quad . \quad . \quad (10)$$

where "end effects" are again apparent in the last two terms in the right-hand side of (10). Let us now write

$$\boldsymbol{\xi}'\mathbf{A}^{-1}\boldsymbol{\xi}=\frac{1+\lambda^2}{1-\lambda^2}[\mathrm{S}+\mathrm{U}(\lambda)],\ \text{where }\mathrm{S}=\sum_1^n\xi_r^2\ ;\ \text{and also }\mathrm{S}_k=\sum_1^{n-k}\xi_r\xi_{r+k}.$$

Equation (6) of § 4, which determines $\hat{\lambda}$, may then be written

$$\frac{d}{d\lambda}\left\{\frac{\mathrm{S}+2\mathrm{U}(\lambda)}{1-\lambda^2}\right\}=0,$$

whence

$$\hat{\lambda}=-\frac{1}{\mathrm{S}}\{(1-\hat{\lambda}^2)\mathrm{U}'(\hat{\lambda})+2\hat{\lambda}\mathrm{U}(\hat{\lambda})\}. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where $\mathrm{U}'=d\mathrm{U}/d\lambda$. Note the expectation value of S_k is zero for $k>1$; the first approximation to the solution of (11) is obtained by neglecting such terms, as well as those having λ^k as factor, with $k>2$:

$$\frac{\hat{\lambda}_1}{1+\hat{\lambda}_1^2}=-\frac{\mathrm{S}_1}{\mathrm{S}}=\alpha_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11\,a)$$

Substitution in (10) gives the first approximation to $\hat{v}$:

$$n\hat{v}_1=\mathrm{S}\left(\frac{1+\hat{\lambda}_1^2}{1-\hat{\lambda}_1^2}\right)\left(1+2\hat{\lambda}_1\frac{\mathrm{S}_1}{\mathrm{S}}\right)=\mathrm{S}=\sum_{r=1}^{n}\delta\psi_r^2-\psi^2/n. \quad . \quad . \quad . \quad (12)$$

Using the relation $\alpha=w/v$, we obtain

$$n\hat{w}_1=-\sum_{r=1}^{n-1}\delta\psi_r\delta\psi_{r+1}+\psi^2/n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Finally, since $s^2t=n(v-2w)$, the first approximation to the maximum likelihood estimate of the mean-square scattering deviation is

$$\hat{s}_1^2t=\sum_{r=1}^{n}\delta\psi_r^2+2\sum_{r=1}^{n-1}\delta\psi_r\delta\psi_{r+1}-3\psi^2/n=\theta_s^2 \quad . \quad . \quad . \quad . \quad . \quad (14)$$

(we shall use θ_s to denote the observed r.m.s. scattering deviation in the whole track). Note that this estimate of s^2 *is independent of the smoothing factor* q: this is a desirable feature in practice, since the value of q will depend to a certain extent on the idiosyncrasies of individual observers.

Equation (11) is in the right form for obtaining higher order approximations to $\hat{\lambda}$ by iteration: the kth iterate would be

$$\lambda_k=-\frac{1}{\mathrm{S}}\{(1-\hat{\lambda}_{k-1}^2)\mathrm{U}'(\hat{\lambda}_{k-1})+2\hat{\lambda}_{k-1}\mathrm{U}(\hat{\lambda}_{k-1})\}. \quad . \quad . \quad . \quad (15)$$

We can find out whether higher order approximations are worth obtaining by calculating the efficiencies of the first order approximations above, as defined at the end of the last paragraph. Their variances (standard errors squared) may be calculated by the methods of § 2 and § 3 : the algebra is somewhat tedious and need not be reproduced; neglecting some terms whose ratios to the main terms is of order $1/n$,

$$\left.\begin{aligned} \epsilon^2(\hat{\omega}_1 t)&=E(\psi-\overline{\psi})^2=s^2t+2w, \\ \epsilon^2(\hat{s}_1^2 t)&=E(\theta_s^2-\overline{\theta_s^2})^2=\frac{2}{n}\{(s^2t+2nw)^2+2(s^2t)^2\}. \end{aligned}\right\} \quad . \quad . \quad (16)$$

The determinants B_n in (8), (9) of § 4 and A_n in (1) of this paragraph are related by

$$\log B_n=n\log(v/s^2h)+\log A_n=n\log b, \quad . \quad . \quad . \quad . \quad (17)$$

so that, neglecting terms of order $1/n$ and λ^n compared to unity,

$$b=\tfrac{1}{2}\{a+2s_0^2/s^2h+(2a-1+4s_0^2/s^2h)^{\frac{1}{2}}\}, \quad . \quad . \quad . \quad . \quad (17\,a)$$

where $a=\frac{2}{3}(1-\frac{1}{2}q+q^2)$. We therefore obtain the variances of the (exact) maximum likelihood estimates by substituting from (17) in their expression (9) of § 4, taking $\xi=s_0^2/s^2h$: thus

$$\left.\begin{aligned} \epsilon^2(\hat{\omega}t)&=\frac{s^2h}{\mathbf{h}'\mathbf{B}^{-1}\mathbf{h}}t^2=\frac{n^2v}{\sum\limits_{r,s}A_{rs}}=nv\left(1-\frac{2w}{v}\right)=s^2t, \\ \epsilon^2(\hat{s}^2t)&=\frac{2(s^2t)^2}{n}\left\{1-\frac{(db/d\xi)^2}{bd^2b/d\xi^2}\right\}=\frac{2(s^2t)^2}{n}\{1+2(2a-1+4s_0^2/s^2h)^{\frac{1}{2}}\}. \end{aligned}\right\} . \quad (18)$$

The required efficiencies of the approximate estimates are therefore

$$\left.\begin{aligned} \text{Eff. }(\hat{\omega}_1 t)&=\lim_{t\to\infty}\left\{\frac{s^2t}{s^2t+2w}\right\}=1, \\ \text{Eff. }(\hat{s}_1^2 t)&=\frac{1+2(2a-1+4s_0^2/s^2h)^{\frac{1}{2}}}{2+(a+2s_0^2/s^2h)^2}. \end{aligned}\right\} \quad . \quad . \quad . \quad (19)$$

The efficiency of $\hat{s}_1^2 t$, taking $a=1$, has been plotted in fig. 1 against the ratio $\rho=s_0/s\sqrt{h}$: it is $\sim$90 per cent near optimum conditions, where $\rho\sim\frac{1}{2}$*. We conclude therefore that *one can be content as a rule with the first approximation to the maximum likelihood estimates of* ω *and* s^2, as given in equations (9 *a*) and (14) of this paragraph; in what follows we shall no longer distinguish between the approximate and the exact estimates.

The fact that the estimate of the magnetic deviation is simply the total deviation in the track may seem surprising at first sight, but a moment's reflection will show that it must be so. It is clearly the best estimate when $s_0=0$ and $q=1$, since the observed deviations are then all independent, and its variance is then s^2t. It follows that no improved estimate could possibly reduce the error, as given in equation (16), by a factor greater than $1+2w/s^2t$, which is of order $1+1/2n$ at optimum cell-length.

---

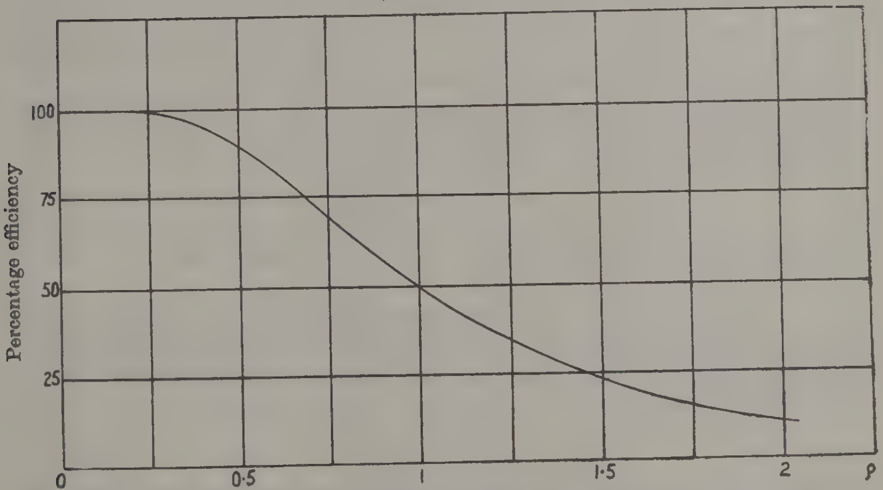
* Fig. 1 can still be used for $a\neq 1$ if we take $[\rho^2-\frac{1}{2}(1-a)]^{\frac{1}{2}}$ as abscissæ; *e. g.* with $a=5/8$ (which is its minimum value) and $\rho=\frac{1}{2}$ the efficiency is >99 per cent.

§ 6. ESTIMATES OF MOMENTUM. OPTIMUM CELL-LENGTH.

It is proposed in this section to compare estimates of momentum obtained from magnetic and scattering deviations under optimum conditions, using the angular difference method. Before doing so, however, we may note the following application for joint measurements of the two deviations : the ratio

$$\gamma_0=\left|\frac{\omega}{s}\right|=\left|\frac{300eH}{\nu}\right| \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Fig. 1.



The ordinates represent the percentage efficiency of the approximate estimates of mean-square scattering deviations, plotted against the ratio $\rho=s_0/s\sqrt{h}$ of the r.m.s. spurious to scattering deviations.

of the magnitudes of the magnetic to the r.m.s. scattering deviations per unit track, ω and s, is *independent of momentum and length of track* (equations (1) and (3) of § 2). Since e and H are known, *it will therefore be possible to combine measurements from a large number of tracks in estimating* γ_0, *and hence verify the theoretical value for the scattering constant* ν, *or alternatively determine it experimentally.* For e=charge of the electron, $\nu=8{\cdot}48\times10^6$ and H=34,000 gauss, $\gamma_0=1{\cdot}20$ cm.$^{-\frac{1}{2}}$. An estimate γ of γ_0 from the data in a single track is clearly, from the results of § 5,

$$\gamma=\frac{|\hat{\omega}|}{\hat{s}}=\frac{|\psi|}{\theta_s\sqrt{(t)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1\,a)$$

An approximate expression for the standard error of this estimate in terms of the standard errors of $\hat{\omega}$ and $\hat{s}$ (known from equations (16) of § 5) is easily obtained by well-known methods :

$$\epsilon^2(\gamma) \simeq \frac{\gamma_0^2}{t}\left\{\frac{\epsilon^2(\hat{\omega}t)}{(\omega t)^2}+\frac{\epsilon^2(\hat{s}^2t)}{4(s^2t)^2}\right\}=\frac{1}{t}+\frac{\gamma_0^2}{2n}\left\{2+\left(1+\frac{2n\omega}{s^2t}\right)^2+\frac{4n\omega}{(\omega t)^2}\right\}. \quad (2)$$

When n is large and the spurious scattering small, $\epsilon(\gamma) \simeq 1/\sqrt{(t)}$. An improved estimate γ_N of γ_0 may be obtained in the usual manner from the data in N tracks (with same e, H and ν) by taking a weighted average of the estimates γ_k from each track :

$$\gamma_N = \sum_{k=1}^{N} \frac{\gamma_k}{\epsilon^2(\gamma_k)} \Big/ \sum_{k=1}^{N} \frac{1}{\epsilon^2(\gamma_k)}. \quad (3)$$

The standard error of γ_N is approximately

$$\epsilon^2(\gamma_N) = N \Big/ \sum_{k=1}^{N} \frac{1}{\epsilon^2(\gamma_k)}. \quad (4)$$

It is equally possible, of course, to combine the data from tracks with different e_k and H_k by weighting each γ_k inversely as e_kH_k.

We now come to the estimation of momentum. An estimate from the magnetic deviation, denoted by p_m, is clearly

$$p_m = \frac{300eH}{\hat{\omega}} = \frac{300eHt}{\psi}. \quad (5)$$

Its relative error is approximately

$$\frac{\epsilon(p_m)}{p} \simeq \frac{\epsilon(\hat{\omega}t)}{\hat{\omega}t} \simeq \frac{1}{\gamma_0\sqrt{(t)}}\left(1+\frac{2\omega}{s^2t}\right)^{\frac{1}{2}}. \quad (6)$$

Provided that t is large enough and the spurious scattering small enough for $2\omega/s^2t$ to be $\ll 1$, the product of this relative error by $\sqrt{(t)}$,

$$\eta(p_m) = \sqrt{(t)}\,\epsilon(p_m)/p \simeq 1/\gamma_0 \quad (7)$$

is approximately equal to $1/\gamma_0$, and hence independent of momentum and length of track and inversely proportional to H.

The scattering estimate of momentum may be written

$$p_s = \frac{\nu}{\hat{s}} \doteq \frac{\nu\sqrt{(t)}}{\theta_s}; \quad (8)$$

its relative error is approximately

$$\frac{\epsilon(p_s)}{p} \simeq \frac{\epsilon(s^2t)}{2s^2t} \simeq \frac{1}{\sqrt{(2n)}}\left[\left(1+\frac{2n\omega}{s^2t}\right)^2+2\right]^{\frac{1}{2}}. \quad (9)$$

In order to compare these two estimates, we shall have to find out how changes in cell-length affect the error on p_s. The standard error of $\hat{s}^2t$ (equation (16) of § 5) may be written as follows :

$$\epsilon^2(\hat{s}^2t) = 2ht\left\{\left(as^2+\frac{2s_0^2}{h}\right)^2+2s^4\right\}, \quad (10)$$

where $a=\frac{2}{3}(1-\frac{1}{2}q+q^2)$. A decrease in the magnitude of h must, on the one hand, tend to decrease this error, because it increases the number $n=t/h$ of observations ; on the other hand, we may expect it to increase the spurious deviation s_0, and hence increase the error. There must, therefore, be an *optimum value*, say h_0, of h, dependent on the way s_0 varies with h, which will minimize $\epsilon(\hat{s}^2t)$. A stationary value is obtained by solving $d\epsilon^2/dh=0$; this leads to

$$\frac{s_0^2}{h_0}-\frac{d(s_0^2)}{dh_0}=\left(\frac{d(s_0^2)}{dh_0}\right)^2+as^2\frac{d(s_0^2)}{dh_0}+\frac{2+a}{4}s^4. \quad . \quad . \quad . \quad (11)$$

It is easily shown that this gives a minimum, provided that

$$\frac{d^2(s_0^2)}{dh_0^2}>-\frac{2}{h_0}\left(\frac{d(s_0^2)}{dh_0}-\frac{s_0^2}{h_0}\right)\Big/\left(s^2-\frac{2s_0^2}{h_0}\right).$$

There is experimental evidence that the dependence of s_0 on h is in the form of a power law (*cf.* Dilworth, *et. al.* 1950),

$$s_0^2=\alpha^2h^{-\mathrm{K}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

with α roughly independent of momentum, and $\mathrm{K}\simeq 3/2$. Substituting in (11), we find that the ratio ρ of spurious to scattering deviations at optimum cell-length,

$$\rho^2=\frac{s_0^2}{s^2h_0}=\frac{a\mathrm{K}}{2(1+2\mathrm{K})}\left\{\left[1+\frac{(2+a)(1+2\mathrm{K})}{a^2\mathrm{K}^2}\right]^{\frac{1}{2}}-1\right\}, \quad . \quad . \quad (13)$$

is determined by the values of K and a (a can range from $\frac{5}{8}$ to 1 ; *cf.* § 3). Table 1 shows that ρ is not very sensitive to variations of either a or K. The explicit expression for the optimum cell-length in terms of momemtum is

$$h_0=\left(\frac{\alpha}{\rho\nu}p\right)^{\frac{2}{1+\mathrm{K}}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

TABLE I.

K	1	1	3/2	3/2	3	3
a	1	5/8	1	5/8	1	5/8
ρ	0·60	0·61	0·53	0·55	0·42	0·45

Substituting this in (10) and (9), one finds for $\sqrt{(t)}$ times the relative error on p_s *at optimum cell-length*

$$\eta(p_s)=\frac{\sqrt{(t)}\epsilon(p_s)}{p}=[1+\tfrac{1}{2}(a+2\rho^2)]^{\frac{1}{2}}\left(\frac{\alpha}{\rho\nu}p\right)^{\frac{1}{1+\mathrm{K}}}. \quad . \quad . \quad . \quad (15)$$

Both h_0 and $\eta(p_s)$ increase with p. Since there are practical limits to the cell-length when one uses a microscope, there will be a range of values of p outside which one will have to use respectively the maximum and minimum practicable values of h. For such values of p, h is fixed and $\eta(p_s)$ becomes

$$\eta(p_s)=h^{\frac{1}{2}}\left[1+\tfrac{1}{2}\left(1+2h^{-(K+1)}\left(\frac{\alpha}{\nu}p\right)^2\right)^2\right]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

With $\alpha=4{\cdot}25\times10^5$, $K=3/2$*, $\nu=8{\cdot}48\times10^6$ and $a=1$

$$h_0\,(\text{microns})=(p\ \text{MeV/c.})^{4/5},$$

$$\eta(p_s)\ \text{cm.}^{\frac{1}{2}}=1{\cdot}5\times10^{-2}\ (p\ \text{MeV/c.})^{2/5};\quad (h=h_0),$$

and the optimum cell-length range with 25 microns $\leqslant h \leqslant$ 400 microns, is 60 MeV/c. $\leqslant p \leqslant$ 2000 MeV/c.

We saw before that $\eta(p_m)$ for the magnetic estimate of p is very approximately independent of p : hence for every value of H there will be a critical value of the momentum below which the scattering determination of p is better than the magnetic one. In the optimum cell-length range, and with the values of α, ν, K and a above, this critical value is

$$p\ \text{MeV/c.}=(1{\cdot}76\times10^6/\text{H gauss})^{5/2}.$$

§ 7. DETERMINATION OF THE SIGNS OF PARTICLES. ESTIMATION OF THE RATIO OF POSITIVE TO NEGATIVE PARTICLES.

It was seen in the last section that scattering determinations of momentum are more accurate than magnetic ones for all momenta smaller than a critical value. There is, however, one important item of information which can be inferred only from the magnetic deviation, and that is information regarding the *sign* of the particle ; such information is available provided we know the direction of travel of the particle.

The magnetic deflection is estimated, as seen in § 5, by the total deviation ψ of the track. What we require, therefore, is the probability, given this deviation, and assuming the direction of travel is known, that the particle should be positive or negative. Let us introduce a charge variable τ, with $\tau=1$ for positive, $=-1$ for negative, particles. Let $p(\tau)$ be the *prior* sign probability of the particle ; we shall also write $q=p(1)/p(-1)$ for the ratio of positive to negative particles (not to be confused with the "smoothing" factor of § 3). We shall assume at first that q is known, either for primaries or for star secondaries ; we shall see later how q can be *estimated* from the statistics of particles. The probability of a deviation ψ *given* the sign of the particle may be written, if we take ω to mean the *magnitude* of the magnetic deviation per unit track,

$$\text{P}(\psi|\tau)=\frac{1}{(2\pi s^2t)^{\frac{1}{2}}}\,e^{-(\psi-\tau\omega t)^2/2s^2t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

---

* Based on preliminary results ; see Dilworth *et. al.* 1950, for improved results and discussion.

Hence the joint probability of ψ and τ is

$$\mathrm{P}(\psi, \tau)=\mathrm{P}(\psi \mid \tau)p(\tau), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

the probability of ψ irrespective of the sign of the particle is

$$\mathrm{P}(\psi)=p(1)\mathrm{P}(\psi \mid 1)+p(-1)\mathrm{P}(\psi \mid -1). \quad . \quad . \quad . \quad . \qquad (3)$$

Finally, the probability of τ given ψ is

$$\mathrm{P}(\tau \mid \psi)=\frac{\mathrm{P}(\psi, \tau)}{\mathrm{P}(\psi)}=\left[1+\frac{p(-\tau)}{p(\tau)}\, e^{-2\tau\omega\psi/s^2}\right]^{-1}. \quad . \quad . \quad . \quad . \qquad (4)$$

(We may note at this point that (4) is essentially an application of Bayes' theorem, which is indubitably valid here since the prior probabilities $p(\tau)$ can be observed.) Equation (4) expresses the maximum information a track can yield regarding sign : this can be shown in the following way. All the information obtained from the track is contained in the set of observed deviations $\delta\psi_r$ $(r=1, 2, \ldots, n)$. Let $\mathbf{z}$ stand as before (§ 3) for this set of observations. The considerations above may be repeated word for word in terms of the probabilities of $\mathbf{z}$ and τ instead of ψ and τ, yielding

$$\mathrm{P}(\tau \mid \mathbf{z})=\frac{\mathrm{P}(\mathbf{z}, \tau)}{\mathrm{P}(\mathbf{z})}=\left[\frac{p(1)}{p(\tau)}\,\frac{\mathrm{P}(\mathbf{z} \mid 1)}{\mathrm{P}(\mathbf{z} \mid \tau)}+\frac{p(-1)}{p(\tau)}\,\frac{\mathrm{P}(\mathbf{z} \mid -1)}{\mathrm{P}(\mathbf{z} \mid \tau)}\right]^{-1}. \quad . \quad . \qquad (4\,a)$$

Substituting expression (8) of § 2 for $\mathrm{P}(\mathbf{z} \mid \tau)$ in (4 *a*), we get

$$\mathrm{P}(\tau \mid \mathbf{z})=\left[1+\frac{p(-\tau)}{p(\tau)}\exp.\,(-2\tau\omega\mathbf{h}'\mathbf{\Lambda}^{-1}\mathbf{z})\right]^{-1},$$

or, in terms of the maximum likelihood estimates of ω and its variance (expressions (4) and (5) of § 4),

$$\mathrm{P}(\tau \mid \mathbf{z})=\left[1+\frac{p(-\tau)}{p(\tau)}\exp\,(-2\tau\omega\hat{\omega}/\epsilon^2(\hat{\omega}))\right]^{-1}, \quad . \quad . \quad . \qquad (4\,b)$$

which reduces to (4) if one substitutes expressions (9 *a*) and (18) of § 5 for $\hat{\omega}$ and $\epsilon^2(\hat{\omega})$.

Expression (4) is exact if we know s and ω (*i. e.* if we know the momentum of the particle). If not, then $\omega/s=\gamma_0$, which is a constant depending only on the field and the composition of the emulsion (*cf.* § 6), and

$$\frac{\psi}{s}=\frac{\psi\sqrt{(t)}}{\theta_s}=\gamma t$$

can be estimated from the track, where γ is now taken as *signed*. Hence, an estimate of (4) is

$$\mathrm{P}(\tau \mid \psi)=\left[1+\frac{p(-\tau)}{p(\tau)}\, e^{-2\tau\gamma\gamma_0 t}\right]^{-1}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

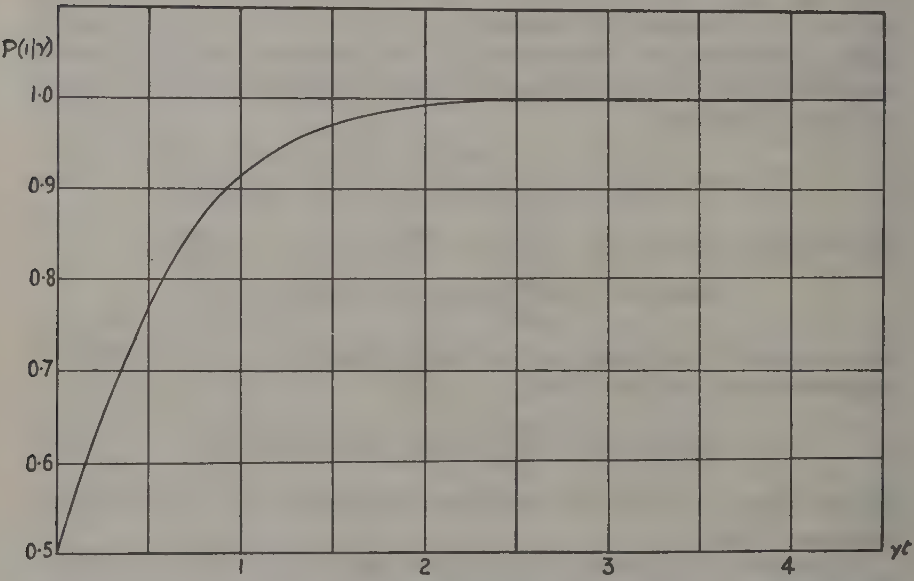
This estimate becomes dubious for high energy particles, such that the error on the estimate $\hat{s}$ is large. In this case we may consider the problem from a different angle. If the number of cells n is large, the distribution of the observed γ given τ is approximately Gaussian, with mean $\tau\gamma_0$.

The probability $P(\tau|\gamma)$ of sign *given* γ may then be obtained by substituting $1/\epsilon^2(\gamma)$ (*cf.* equation (2) of § 6) for t in (5). In fig. 2, $P(1|\gamma)$ is plotted against γt for $\gamma_0=1{\cdot}20$ and $q=1$.

Another problem arising in connection with sign determination is the estimation of the ratio q of positive to negative particles from a set of n tracks of a certain type : *e. g.* primaries in a certain range of momenta, or secondaries from certain types of stars. The probability of a total deviation ψ_r in the rth track, irrespective of sign, may be written

$$P_r(\psi_r)=(2\pi s_r^2 t_r)^{-\frac{1}{2}} e^{-(\psi_r^2+\omega_r^2 t_r^2)/2 s_r^2 t_r}\left[p_+ e^{\omega_r\psi_r/s_r^2}+(1-p_+)e^{-\omega_r\psi_r/s_r^2}\right],$$
$$(r=1, 2, \ldots, n) \quad . \quad . \quad . \quad . \quad (6)$$

Fig. 2.



A plot of the probability $P(1|\gamma)$ of the particle being positive as a function of $\gamma t=\psi\sqrt{(t)}/\theta_s$ for $\gamma_0=1{\cdot}20$.

where $p_+=p(+1)$, t_r, s_r are respectively the length and r.m.s. scattering deviation per unit track for the rth track. Since the deviations in the n tracks are independent, the maximum likelihood estimate of p_+ is obtained by maximizing

$$L=\sum_{r=1}^{n}\log P_r(\psi_r),$$

i. e. by solving

$$\frac{\partial L}{\partial p_+}=\sum_{r=1}^{n}\frac{e^{\omega_r\psi_r/s_r^2}-e^{-\omega_r\psi_r/s_r^2}}{p_+e^{\omega_r\psi_r/s_r^2}+(1-p_+)e^{-\omega_r\psi_r/s_r^2}}=0.$$

Hence the maximum likelihood estimate of $q=p_+/(1-p_+)$ in terms of the n observed deviations is

$$\hat{q}=\frac{\sum_1^n\left[1+\frac{1}{\hat{q}}e^{-2\omega_r\psi_r/s_r^2}\right]^{-1}}{\sum_1^n[1+\hat{q}e^{2\omega_r\psi_r/s_r^2}]^{-1}}\simeq\frac{\sum_1^n \mathrm{P}(1\,|\,\gamma_r;\,\hat{q})}{\sum_1^n \mathrm{P}(-1\,|\,\gamma_r;\,\hat{q})}, \quad . \quad . \quad . \quad . \quad (7)$$

where γ_r is the value of γ observed for the rth track. Equation (7) involves $\hat{q}$ on both sides and must be solved by a method of successive approximation. The following iteration method may be used: take $q_0=1$ as the zero order approximation; then the kth order approximation q_k is obtained from the $(k-1)$th by

$$q_k=\frac{\sum_1^n \mathrm{P}(1\,|\,\gamma_r;\,q_{k-1})}{\sum_1^n \mathrm{P}(-1\,|\,\gamma_r;\,q_{k-1})} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If the exact solution $\hat{q}>1$ then it is easily seen that (*a*) the sequence q_k is an increasing one, and (*b*) that, for every k, $q_k\leqslant\hat{q}$; conversely, if $\hat{q}<1$, then the sequence q_k is a decreasing one, and $q_k\geqslant\hat{q}$ for every k. Hence, in both cases q_k converges to $\hat{q}$ as k increases.

ACKNOWLEDGMENTS.

It is a pleasure to acknowledge my indebtedness to Dr. G. D. Rochester and his group, and in particular Mme F. Lévy and Mr. S. J. Goldsack, for suggesting this problem and keeping me informed about their experimental work; to Professor M. S. Bartlett for his valuable suggestions; to M. Yves Goldschmidt-Clermont for stimulating discussion and correspondence. Mr. Goldsack suggested some of the results in § 7.

REFERENCES.

CRAMÉR, 1946, *Mathematical Methods of Statistics*, pp. 498–506.
DILWORTH, GOLDSACK, GOLDSCHMIDT-CLERMONT and LÉVY, 1950, *Phil. Mag.*, **41**, 1032.
FOWLER, 1950, *Phil. Mag.*, **41**, 169.
GOLDSCHMIDT-CLERMONT, KING, MUIRHEAD and RITSON, 1948, *Proc. Phys. Soc.*, **61**, 183.
GOLDSCHMIDT-CLERMONT, 1950, *Bulletin du Centré de Physique Nucléaire de l'Université de Libre Bruxelles*, No. 14.
GOUDSMIT and SAUNDERSON, 1940, *Phys. Rev.*, **57**, 24, 552.
KENDALL, 1946, *The Advanced Theory of Statistics*, Vol. 2, pp. 12–42.
KOOPMANS, 1942, *Ann. Math. Stats.*, **13**, 27.
LEWIS, 1950, *Phys. Rev.*, **78**, 526.
MOLIÈRE, 1947, *Z. für Naturforschung*, **2a**, 133; 1948, *Ibid.*, **3a**, 78.
MOYAL, 1949, *J. Roy. Statist. Soc. Suppl.*, **11**, 150.
ROSSI and GREISEN, 1941, *Rev. Mod. Phys.*, **13**, 240.
SNYDER and SCOTT, 1949, *Phys. Rev.*, **76**, 220.
WILLIAMS, 1939, *Proc. Roy. Soc.* A, **169**, 531.

XCVI. CORRESPONDENCE

The Velocity of Second Sound Below 1° *K.*

By K. R. Atkins and D. V. Osborne,
Royal Society Mond Lab., Cambridge *.

[Received August 15, 1950.]

Peshkov (1948), Maurer and Herlin (1949) and Pellam and Scott (1949) have shown that, as predicted by Landau (1941), the velocity of second sound in liquid helium II increases as the temperature is decreased below 1° K. We have extended these measurements to lower temperatures and have found that the velocity rises to 152 ± 5 metres/sec. at 0·1° K. Subject to the uncertainties which will be discussed below, this agrees with Landau's further prediction that, at 0° K., the velocity of second sound should be equal to the velocity of first sound divided by $\sqrt{3}$.

The low temperatures were produced by demagnetizing a mixture of 2 g. of liquid helium and 30 g. of ferric alum from a field of 10,000 gauss. The mixture was contained in a small metal Dewar vessel and was pumped by a diffusion pump of speed 7 litres/sec. through a constriction of diameter 0·1 mm. Owing to certain technical difficulties, this cryostat has not yet behaved ideally, and the warming times were only of the order of 60 sec. However, this was sufficient to enable the velocity measurements to be made and the preliminary results described below establish the main characteristics of the variation of velocity with temperature. Temperatures after demagnetization were deduced from the magnetic susceptibility of the ferric alum, measured by means of a mutual inductance and a ballistic galvanometer. The primary current of the mutual inductance was automatically reversed every three sec. and the throws of the ballistic galvanometer were photographically recorded.

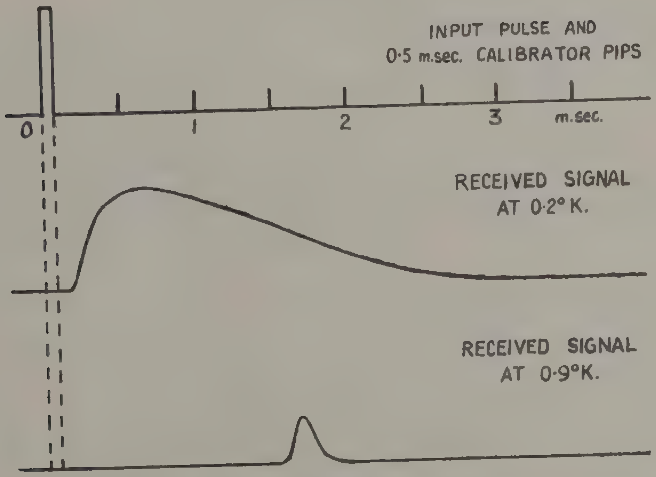
The second sound measurements used a pulse technique, in which a D.C. pulse of duration 100 μ sec. was fed into a constantan wire heater. The resulting second sound pulse was propagated along a glass tube of length 3·40 cm. and internal diameter 0·5 cm. embedded in the ferric alum, and was received by a phosphor-bronze resistance thermometer. The time of flight was measured by displaying the received signals on a cathode-ray tube, which was photographed with a cine camera during the 60 sec. immediately following the demagnetization. Fig. 1 has been built up from tracings of actual photographs. In agreement with the observations of Pellam and Scott (1949), the received pulse is seen to be considerably distorted and lengthened at these low temperatures. This distortion can be explained in terms of dissipative losses during propagation, and a full analysis will be published shortly. No detectable change in velocity was caused by altering the pulse amplitude, so that the shock-wave effects discovered by one of us (Osborne, to be published shortly) do not seem to be relevant to the present results.

---

* Communicated by the Authors.

Fig. 2 presents our results for the variation of velocity with temperature, together with previous results at higher temperatures. Owing to the large heat influx, it is uncertain how closely the temperature of the liquid followed that of the ferric alum, and so the temperatures must be considered as only approximate, especially above 0·3° K. when the warming up was particularly rapid. Landau (1941) predicts that the velocity of second sound (u_2) at 0° K. should be $1/\sqrt{3}$ times the velocity of ordinary sound (u_1). The existing measurements of u_1 (Findlay *et al.*, 1938, Pellam and Squire, 1947) do not extend below 1·5° K. and the extrapolation to 0° K. is rather uncertain. A linear extrapolation gives, at 0° K, $u_1=272$ metres/sec. and $u_1/\sqrt{3}=157$ metres/sec. Therefore, apart from the uncertainty of this extrapolation and the possibility that u_2 changes appreciably between 0·1° K and 0° K., the relation $u_2=u_1/\sqrt{3}$ is well borne out by fig. 2. In Landau's theory the falling away of u_2 from this initial value is due to rotons, and fig. 2 therefore suggests that rotons are becoming important at about 0·1° K. and are predominant above 0·5° K. For this reason it seems unlikely that the specific heat of liquid helium will obey a T^3 law above 0·1° K.

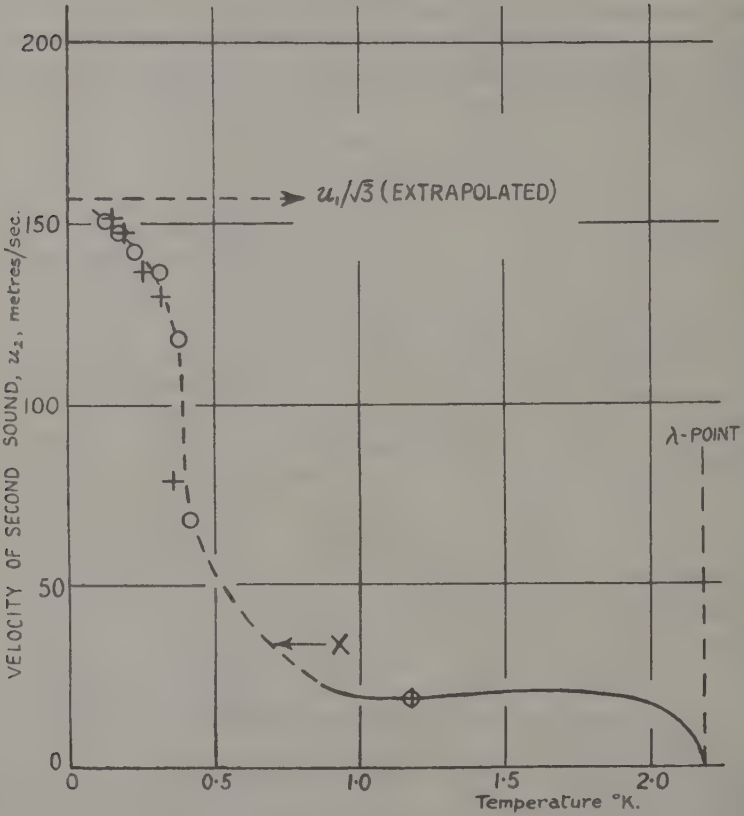
Fig. 1.



Tracings taken from photographs of the oscilloscope screen.

The two aspects of Landau's theory which are mainly responsible for the particular result $u_2=u_1/\sqrt{3}$ are (1) his method of calculating the density of the normal component, ρ_n, and (2) his fundamental assumption that the liquid can be treated as a continuum, in much the same way as a solid is treated in the Debye theory of specific heats, and that the thermal motion can therefore be built up from phonons and rotons. Near 0° K. only phonons are important and, since these are just ordinary sound waves, it is easy to see how u_1 becomes involved in the formula for u_2. The alternative theory developed by London (1938) and Tisza (1947), attaching less importance to the interatomic forces, builds up the wave function of the liquid as a whole from the wave functions of individual atoms, and then makes use of those special properties of Bose–Einstein statistics which lead to a condensation of atoms into the lowest state. Temperley (to be published shortly) has shown that if ρ_n/ρ is equated to the proportion of

Fig. 2.



Variation of velocity with temperature.

○ + The present results. Two separate demagnetisations.
—— Results of Peshkov and of Maurer and Herlin at higher temperatures.
× Highest velocity reached by Pellam and Scott. Temperature unspecified.

atoms not in the lowest state, then this type of theory always leads to $u_2=0$ at 0° K. If, however, ρ_n/ρ is evaluated by a method similar to Landau's, this conclusion remains true for the Bose–Einstein gas, but probably does not hold when interactions between atoms are taken into account. It is conceivable, therefore, that a Bose–Einstein type of theory making due allowance for interatomic forces would give a value of u_2 tending at 0° K. to a finite value related, presumably, to the average velocity of zero-point motion, which has the same order of magnitude as u_1. Nevertheless, until such a theory has been developed and has satisfied the exacting requirement that $u_2=u_1/\sqrt{3}$ at 0° K., the present experimental results must be considered strongly in favour of Landau's method of approach.

We should like to express our indebtedness to Dr. D. Shoenberg for encouraging this research, to Dr. J. Ashmead for technical advice, and to Mr. H. N. V. Temperley for showing us his results before publication.

REFERENCES.

FINDLAY, J. C., PITT, A., GRAYSON SMITH, H. and WILHELM. J. O., 1938, *Phys. Rev.*, **54**, 506.
HERLIN, M. A., and MAURER, R. D., 1949, *Phys. Rev.*, **76**, 948.
LANDAU, L., 1941, *J. Phys., U.S.S.R.*, **5**, 71.
LONDON, F., 1938, *Phys. Rev.*, **54**, 947.
PELLAM, J. R., and SCOTT, R. B., 1949, *Phys. Rev.*, **76**, 869.
PELLAM, J. R., and SQUIRE, C. F., 1947, *Phys. Rev.*, **72**, 1245.
PESHKOV, V., 1948, *J. Phys., U.S.S.R.*, **18**, 951.
TISZA, L., 1947, *Phys. Rev.*, **72**, 838.

The Velocity of Flow of Helium II in the Creeping Film.

By L. C. JACKSON and D. G. HENSHAW*,
H. H. Wills Physical Laboratory, University of Bristol.

[Received August 30, 1950.]

IN their paper on the measurement of creep rates and the thickness of the helium II film Daunt and Mendelssohn (1939) combined the two results to calculate the average velocity of flow of the helium in the film. Their value (about 20 cm./sec. at 1·5° K.) can, however, only have an order of magnitude significance as the thickness determination was made on a stationary film and was an average value over a considerable range of heights above the liquid level and the determinations of the volume of helium transported per sec. were of necessity made on the moving film.

We have, therefore, undertaken the simultaneous measurement with the same apparatus of the rate of creep and the thickness of the creeping film, using for the latter the optical method of Jackson and Burge (1949).

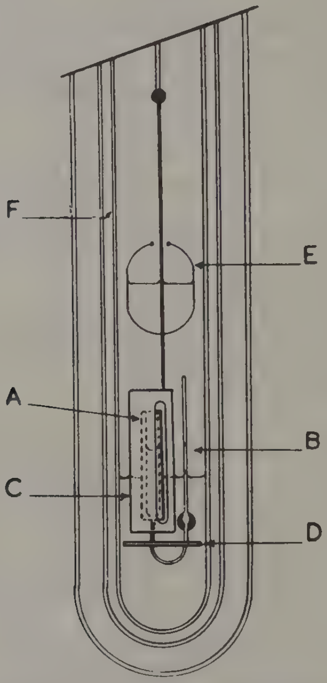
---

[*] On leave of absence from the National Research Council of Canada.

In addition to providing the mean velocity of flow of the helium II at various temperatures, this work has brought to light an interesting feature of the flow itself which, in our opinion, warrants a preliminary report at this time.

The beaker A (fig. 1) is of stainless steel, 4 mm. internal diameter, 1 mm. wall thickness. A polished 2 mm. wide flat on the outer surface of the beaker is used for the optical measurements of the thickness of the helium film, while a glass capillary B, attached to the beaker by means of a copper-nickel tube and a kovar-glass seal is used for the liquid level measurements. The beaker is mounted in a brass radiation shield C and is supported at the bottom by means of the copper-nickel capillary tube. A disc D is attached to the capillary tube to maintain the beaker central in the dewar. The assembly is then suspended from a glass and tungsten support to which is attached a glass beaker E, which holds liquid helium during the experiment and which provides additional thermal protection to the stainless steel beaker. The apparatus is mounted in the dewar Γ which contains the liquid helium, the latter being protected by liquid hydrogen and liquid nitrogen in the outer dewars.

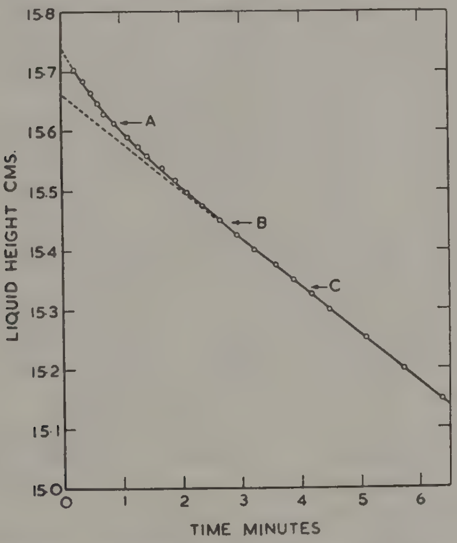
Fig. 1.



Apparatus for simultaneous measurement of film thickness and creep rate. A, stainless steel beaker; B, glass capillary tube; C, radiation shield; D, centering disc; E, glass beaker on support; F, helium dewar.

The creep rates observed are of the same order of magnitude as those reported by Mendelssohn and White (1950) for cleaned platinum and considerably in excess of those for glass beakers. Although the stainless steel beaker has mono-layers of barium stearate on the outside, it apparently behaves as a " clean beaker " in the sense of Bowers and Mendelssohn (1949), since the creep rate is practically independent of the height of the inner liquid level except for the first few millimeters, where, as usual, the creep rate is considerably greater (see fig. 2). The accompanying table gives the measured values of the creep rates and the thickness and the deduced value of $\bar{v}$, the helium velocity averaged over the film thickness. The recorded thicknesses are those obtained when the point of observation of the thickness is 1 cm. above the main liquid helium level. The variation of the thickness of the moving film with height will be dealt with in a more detailed communication.

Fig. 2.



Creep curve at 1·89° K. Height of liquid helium II inside beaker v. time. A, drops decreasing in number ; B, very few drops visible ; C, no drops visible.

T° K.	Creep rate cm.^{3}/cm./sec.	Thickness cm.	Velocity $\bar{v}$ cm./sec.
1·1	$16{\cdot}9\times10^{-5}$	$1{\cdot}63\times10^{-6}$	72
1·3	$16{\cdot}9\times10^{-5}$	$1{\cdot}63\times10^{-6}$	72
1·5	$16{\cdot}8\times10^{-5}$	$1{\cdot}66\times10^{-6}$	68
1·7	$16{\cdot}1\times10^{-5}$	$1{\cdot}82\times10^{-6}$	60
1·9	$12{\cdot}8\times10^{-5}$	$1{\cdot}94\times10^{-6}$	29

A very striking phenomenon is observed when the full beaker is withdrawn from the main body of the liquid helium II at the commencement of measurements. The mirror is seen to be covered with a large number of bright specks of light moving down and disappearing into the liquid helium below. Fifty or more may be visible at once immediately after the beaker has been raised. As the level of the liquid in the beaker falls the number of specks decreases steadily, the last ones disappearing in the present case when the liquid is some 3 mm. from the rim of the beaker. The bright specks, which are obviously "drops" of liquid helium, are somewhat elongated in the direction of motion, are 1/10 to 1/3 mm. long and move downwards at a speed of about 1 cm. per sec. It is not yet possible to state anything definite about the dimensions of the drops perpendicular to the mirror. They are certainly much greater than the thickness of the normal creeping film, which is observed when the drops have ceased. The drop thickness is greater than can be measured with the existing mica compensating plate and the drops are too small for it to be possible to say whether they show interference contours. In fig. 2, A marks the position at which the observer reported that the number of drops visible had decreased appreciably, B that at which only a few individual drops were visible on the mirror, and C that for which drops were no longer seen. It will be seen that the region in which the drops are present coincides closely with that in which an anomalously high creep rate is observed. There seems little doubt that the straight part of the creep curve represents a flow of helium characteristic of the continuous film and that in the initial curved part the excess helium is transported in the form of discrete drops. It will be noted that these drops move much more slowly than the film itself and presumably slide over the surface of the film.

REFERENCES.

BOWERS, R., and MENDELSSOHN, K., 1949, *Nature*, **163**, 870.
DAUNT, J. G., and MENDELSSOHN, K., 1939, *Proc. Roy. Soc.* **A, 170, 423.**
JACKSON, L. C., and BURGE, E. J., 1949, *Nature*, **164**, 660.
MENDELSSOHN, K., and WHITE, G., 1950, *Nature*, **166**, 127.

BOOKS RECEIVED

A new scientific publication has been received—'Journal of Atmospheric and Terrestrial Physics', edited by Sir Edward Appleton with the assistance of an eminent Advisory Board. Published by Butterworth-Springer Ltd. (London). Price per vol. (approx. 300 pp.) £3 10*s*. 0*d*.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XCVII. *Deformation Textures in Face-Centred Cubic Metals.*

By E. A. Calnan, B.Sc., and C. J. B. Clews, Ph.D.
National Physical Laboratory, Metallurgy Division*.

[Received July 11, Revised August 8, 1950.]

Introduction.

There have been several previous attempts to explain the deformation behaviour of polycrystalline aggregates in terms of the slip mechanisms operative when single crystals are deformed, but no wholly satisfactory treatment has been devised. In discussing these, reference must first be made to the work of Boas and Schmid (1931), which assumes the operation of the three most favourable slip systems for any particular crystal orientation and derives the end position which is stable for these three systems. This predicts tension and compression textures, for both face-centred and body-centred cubic metals, which are in good agreement with the main features of the observed textures. Some details of the observed textures are, however, not explained. The treatment of Taylor (1938), mathematically more rigorous, has as its basic assumption the perfectly homogeneous deformation of the aggregate. Following the analysis of Mises (1928) that this demands the simultaneous operation of five special glide systems, Taylor has derived the systems and corresponding shears which conform to the principle of virtual work. The magnitudes and directions of the crystal rotations may be deduced from the shears selected, while the critical shear stress for a particular elongation is calculated from the sum of the shears and the single crystal shear-hardening curve. Good agreement with the experimentally determined tensile stress-elongation curve for a polycrystalline aggregate, and the main features of the tension and compression textures for face-centred cubic metals are obtained. The disadvantages of the treatment are first that homogeneous deformation rarely occurs in practice, and secondly that the heavy computation makes its application to metals of other crystal structures virtually impossible.

Pickus and Mathewson (1939) have put forward a theory of the origin of rolling textures in which they assume that unequal participation of the three most favourable glide systems brings the deformation axes to symmetrical end positions, where the resolved shear stresses are equal and the rotations cancel out. The most favourable systems are those for which the product of the resolved shear stress and the cosine of the angle between the glide direction and the direction in which free flow may occur are highest. The relative probability of the end positions are considered to depend on the values of the product function at the end positions. It is shown that all the observed rolling textures in face-centred metals fit in reasonably well with these ideas. The main criticism of this treatment and that of Boas and Schmid appears to be that, although the crystals oriented to correspond to the end positions will no doubt remain there and will slip without rotation, there is no complete explanation of how other crystals reach these positions.

---

* Communicated by the Authors.

One further criticism of all the previous treatments may be noted. Although the simultaneous operation of several slip systems is demanded, there is no explanation of the means by which the most favourable system or systems can apparently withstand resolved shear stresses greater than the critical value until the resolved shear stresses on less favourable systems have reached the critical value and deformation occurs.

MAIN CONCEPTS OF THE PRESENT TREATMENT.

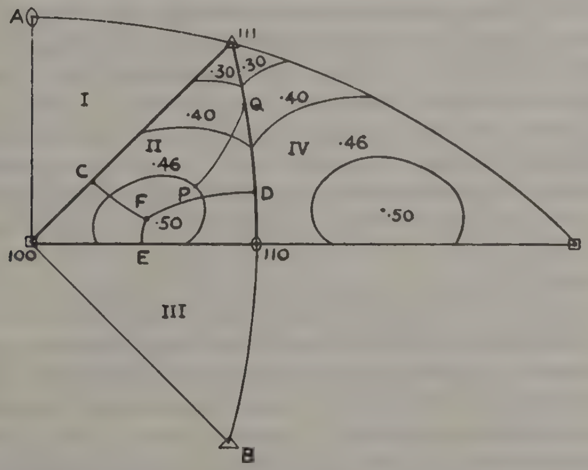
It is an experimental fact that considerable deformation can occur without cracks appearing at the grain boundaries. In the ideal case of homogeneous deformation this demands the operation of five special slip systems, and even allowing for some inhomogeneous deformation, particularly near the grain boundaries, the critical shear stresses must be attained on more than one system at the same time. Moreover, apart from minor differences, for example, the observations on duplex slip in α-brass crystals, Göler and Sachs (1929), the critical shear stresses must be equal. Therefore the constraints imposed by the surrounding grains of the aggregate together with the applied stress must be such as to produce equal resolved shear stresses on the glide systems which operate. That is to say, if a grain deforms by the simultaneous operation of two systems then the resolved shear stress on these systems must be equal, and consequently it will be assumed that the effective stress may be represented by a tensile stress, T_e, so oriented as to give duplex slip, *i.e.* its representative point on the stereographic projection lies on one of the boundaries of the unit triangle. If, however, the grain deforms by the operation of more than two slip systems then T_e must lie at one of the corner points of the triangle, [100], [110], and [111], where there are respectively 8, 4, and 6 equally favourable slip systems.

This concept, although probably an over-simplification of what must be a complex system of stresses, does give a physical picture of the mechanism. Further consideration of the movement of T_e leads to an explanation of the means by which T_e reaches positions favourable for duplex or multiple slip without the resolved shear stress on the most favourable system exceeding the critical value.

The most favourable system for a grain of applied tension axis in unit triangle II of fig. 1 is the direction A on the slip plane of normal B. As the applied stress, T_a, rises the resolved shear stress rises proportionately until it reaches the critical value at which slip would occur if the grain were an isolated single crystal. In general this single slip mechanism does not operate for all the grains in the polycrystalline aggregate, and therefore the constraints due to the surrounding grains must prevent the resolved shear stress from reaching the critical value. Due to these constraints the direction of the effective stress, T_e, moves away from the direction of the applied stress in such a way that T_e resolved on the most favourable system is always below the critical value.

In fig. 1 are shown the contours of equal resolved shear stress for the most favourable system, the numerical values being $\cos\chi \cos\lambda$, where χ and λ are respectively the angles which the slip direction and slip plane normal make with the particular T_a orientation. For orientation P (fig. 1), the critical resolved shear stress C is reached when $T_a = C/0{\cdot}46$, and in order that the critical value shall not be reached as T_a increases, T_e will move away to a point of lower $\cos\chi \cos\lambda$ value, *i.e.*, move down the contour gradient. When T_a has increased such that $T_a = C/0{\cdot}40$, and slip has not occurred, then T_e will lie on the 0·40 contour. The position on the contour will be determined by the stresses exerted by the neighbouring grains, and with a fine grained polycrystalline aggregate the general direction of movement may be taken as the path of quickest descent, *i.e.* perpendicular to the contours. On reaching a boundary of the unit triangle, T_e will travel along it in the direction of decreasing $\cos\chi \cos\lambda$ until it reaches one of the corner points. These are minimum positions, and any further increase in T_a must now result in slip, for in whatever direction T_e attempts to move from the corner point some system will

Fig. 1.



Stereographic projection showing shear stress contours for a face-centred cubic metal.

attain the critical shear stress. Here, then, with the multiple systems available, deformation can occur without immediate misfit at the grain boundaries, and with the symmetrically disposed systems all exactly equivalent there will be no resultant rotation of the crystal.

In some grains T_e may not reach the corner point before slip occurs owing to insufficient lateral stresses, as perhaps in the case of a grain at a free surface. Consequently, rotation will occur in the usual directions of single or duplex slip, the latter taking place if T_e has reached the unit triangle boundary. In addition to those at the surface, the grains most likely to slip before T_e reaches a corner point are those for which the greatest lateral stresses will be required. These, which are the grains most likely to rotate, will thus be those furthest removed from the appropriate corner point, and conversely rotation will be least probable near the corners. In the instant at which slip occurs the lateral stresses are released and T_e returns to T_a or, if rotation takes place, to the new position of T_a. With the cessation of this slip due to strain hardening, the movement of T_e may begin again under the influence of a new environment which may or may not lead to rotation. Since inhomogeneous deformation is implied in this treatment, T_e need not have a unique value throughout the whole of a single grain, and the deformation corresponding to a single movement of T_e may be as small as that associated with the passage of a single dislocation. Thus the deformation process is envisaged as step-like in character rather than as smooth and continuous.

TENSILE STRESS-ELONGATION CURVE.

From the concept of the deformation of a polycrystalline aggregate occurring when T_e in each grain has reached a corner point, the relationship between tensile stress and elongation may be derived from the corresponding single crystal shear hardening curve. First, however, the validity of considering T_e in all the grains as having reached the corner points must be examined, since in this treatment the rotations which lead to deformation textures occur in precisely those grains, or parts of grains, for which this assumption is not true. The fact that the grains do not pull apart at the boundaries must imply that in the boundary regions at least multiple slip occurs. Recent X-ray work on creep deformation in aluminium, Calnan and Burns (1950) has shown the greater complexity of the deformation processes in the regions adjacent to the grain boundaries. In view of the comparatively large size of these boundary regions, in which the lateral stresses must be sufficient to cause considerable displacement of T_e, the lateral stresses elsewhere in the grain are probably not very much less, with the result that T_e will be somewhere near a corner position. It can be shown that the tensile stress for a given elongation is not greatly different for multiple slip deformation (T_e at a corner point) and simple slip deformation with T_e about 5° from the same corner point. Thus an approximation to the polycrystalline curve, within the limits of experimental variations, may be made on the assumption that in all the grains T_e has reached a corner point without vitiating the arguments employed in deriving the deformation textures, namely that in the interior of the grains at least T_e has not reached this point.

On the shear stress contour diagram, fig. 1, the peak position F is joined to the points of maximum shear stress on the boundaries of the unit triangle, C, D and E, by the lines of quickest descent. Since the mean path of T_e is along the line of quickest descent, T_e for all the grains of orientations contained in the area a_{111} between CFD and the [111] moves towards this last direction. For example, T_e for the grain of orientation represented by P moves along its appropriate line for quickest descent until it reaches the [111]–[110] boundary at Q. From the shear stress contours in the adjacent triangle IV, corresponding to a different most favourable slip system, it will be seen that the boundary itself constitutes a line of quickest descent, and thus T_e moves from Q towards [111]. Similarly, T_e for the orientations represented by the area a_{110} between DFE and the [110] moves to the [110], and for those in area a_{100} between CFE and the [100] to the [100]. With an initially random orientation in the polycrystalline aggregate the proportions of grains with these end positions for T_e are given by the ratios of the solid angles to which these areas correspond. It is however more convenient to use the ratios of the areas on the unit triangle, and since this involves an error of only about 1 per cent this method has been adopted.

An elongation e in the direction of the applied stress requires an elongation $e/[\cos^2\phi(1+\sigma)-\sigma]$ in the direction of a given corner point, where ϕ is the angle between the direction of the applied stress and the corner point, and σ is the value of Poisson's ratio chosen to give approximately constant volume over the range of deformation considered. A weighted mean value of $\cos^2\phi$ was calculated for each of the three areas referred to in the previous paragraph, viz. $(\cos^2\phi)_{111}$, $(\cos^2\phi)_{110}$, and $\cos^2\phi)_{100}$. For symmetrical slip it can be shown that the elongation in the direction of the corner point, *i.e.* in the direction of the effective stress for multiple slip, is $s\cos\chi_0\cos\lambda_0$, where s is the sum of the moduli of the shears on the individual operative systems, all of which have the same values of χ_0 and λ_0, the initial angles between the effective stress direction and the slip direction and the slip plane normal respectively. Substituting the appropriate values of χ_0, λ_0 and $(\cos^2\phi)$ in the expression

$$e=s\cos\chi_0\cos\lambda_0[(\cos^2\phi)(1+\sigma)-\sigma],$$

gives the three values of shear s corresponding to the selected value of e. The critical shear stress for each of these s values may be determined from the experimental shear stress—shear strain curve for a single crystal. The effective tensile stress in the direction of the corner point to produce this shear stress is T_e=critical shear stress/$\cos\chi_0\cos\lambda_0$, which in turn implies a tensile stress in the applied stress direction

$$T_a=(\cos\phi)\ .\ \text{critical shear stress}/\cos\chi_0\cos\lambda_0,$$

where $(\cos\phi)$ is the weighted mean value of $\cos\phi$ over the area considered.

TABLE I.

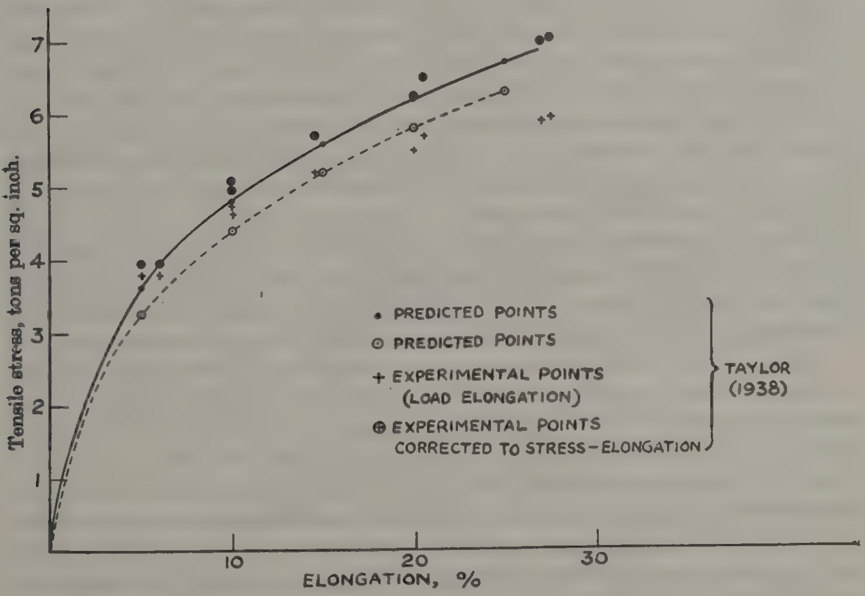
Calculation of the Tensile–Stress–Elongation Curve.

Elongation e%	Weighted mean values $(\cos\phi)_{100}=0{\cdot}96$ $(\cos^2\phi)_{100}=0{\cdot}94$			Weighted mean values $(\cos\phi)_{110}=0{\cdot}97$ $(\cos^2\phi)_{110}=0{\cdot}95$			Weighted mean values $(\cos\phi)_{111}=0{\cdot}92$ $(\cos^2\phi)_{111}=0{\cdot}85$			Weighting factors			Total $\mathrm{T}_{a_{tsi}}$
										(100) 0·172	(110) 0·203	(111) 0·625	
	e'_{100}%	$\mathrm{T}_{e_{100}tsi}$	$\mathrm{T}_{a_{100}tsi}$	e'_{110}%	$\mathrm{T}_{e_{110}tsi}$	$\mathrm{T}_{a_{110}tsi}$	e'_{111}%	$\mathrm{T}_{e_{111}tsi}$	$\mathrm{T}_{a_{111}tsi}$	$\mathrm{T}_{a_{100}tsi}$	$\mathrm{T}_{a_{110}tsi}$	$\mathrm{T}_{a_{111}tsi}$	
5	5·5	2·50	2·41	5·4	2·47	2·41	6·4	4·74	4·36	0·42	0·49	2·72	3·62
10	11·0	3·35	3·23	10·7	3·30	3·22	12·8	6·24	5·74	0·56	0·65	3·59	4·80
15	16·5	3·92	3·77	16·1	3·88	3·78	19·2	7·24	6·66	0·65	0·77	4·16	5·58
20	22·0	4·36	4·20	21·5	4·34	4·23	25·6	8·03	7·39	0·72	0·86	4·62	6·20
25	27·5	4·73	4·56	26·9	4·70	4·58	32·0	8·70	8·00	0·78	0·93	5·00	6·71

$e'=e/(\cos^2\phi)(1+\sigma)-\sigma$; $\sigma=0{\cdot}45$.

The final tensile stress is obtained by adding these three values of T_a weighted in proportion to the areas defined above. The tensile stress has been determined by this method for aluminium up to extensions of about 25 per cent, using the experimental single crystal results of Elam referred to by Taylor (1938). Details of the calculation are listed in Table I., and the resulting polycrystalline stress-elongation curve is shown in fig. 2 together with the experimental points determined by Elam using the same material as that used for the single crystal work. The curve predicted by the Taylor theory is also shown. It will be seen that the curves derived from the Taylor theory and from the present treatment give equally satisfactory agreement with the experimental observations. Although the derivation of a stress-elongation curve must be regarded as an essential part of any deformation theory too much significance should not be attached to the exact agreement of prediction and experiment in view of the various assumptions made. The lack of agreement at higher elongations calls for some comment however. Since the Elam curve is for load-elongation it may be expected to fall below the calculated stress-elongation curves in this region where reduction of cross-sectional area is now no longer negligible. Furthermore, in the calculations random orientation has been assumed throughout, whereas with increasing elongation some slight measure of preferential orientation must occur.

Fig. 2.



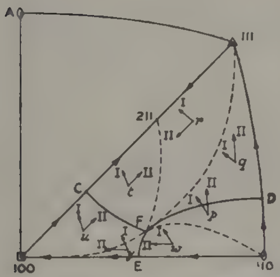
Comparison of predicted tensile stress–elongation curve with Taylor's predicted and experimental curves.

Deformation Texture in Tension.

The development of deformation textures by marked grain rotations implies deformations much more severe than those considered in the stress–strain relationship, where the departure from random orientation due to rotations was not taken into account. It has been proposed above that rotation will occur when the deformation takes place before T_e has reached a corner position. It will be that corresponding to duplex slip if T_e has reached a triangle boundary, and otherwise that corresponding to single crystal slip. In fig. 3 are shown the lines of quickest descent to the corner points, to the [211] direction, and to the maxima on the triangle sides. In addition the rotations for duplex slip are marked with arrowheads along the triangle sides. For an applied stress direction represented by p, the mean path of the effective stress, T_e, is down the contour gradient to reach the [110–111] boundary between [110] and the maximum D. If deformation occurs before T_e reaches this boundary, there is slip on the single most favourable system with consequent rotation along the great circle towards the operative slip direction A. This is shown by the arrow I. If T_e reaches the boundary however, but deformation occurs before it has reached the [110] corner point, rotation corresponding to duplex slip occurs in the direction shown by the arrowhead on the boundary, producing the rotation of p shown by the arrow II. The single and duplex rotations for other typical points q, r, t, u, v and w, are deduced from similar reasoning. As the deformation proceeds, the orientation representing a particular grain (or part of a grain) moves across the unit triangle, and its probable rotations vary according to the particular region in which it is temporarily situated. Thus, with the treatment as oultined so far, it is possible for an orientation such as w to move into the region EF[100], typified by v, then into the region CF[100], and finally into the CF[112] region. One additional feature mentioned in an earlier section will however modify this effect. The grains least likely to rotate are those which require the smallest lateral stresses for T_e to reach the appropriate corner point, *i.e.* those orientations near the corner point. This will apply to the grains near the [111] direction in the DF[111] region, and those near the [100] direction in the EF[100] region. It will not apply to those near [110] in the DF[110] and EF[110] regions or those near [100] in the CF[100] region for the following reason. Consider an orientation similar to p near [110]. Although the effectiveness must travel only a short distance to the [110], any single crystal or duplex rotation which may occur increases the distance from the [110] ,and thus increases the probability of rotation, whereas in the DF[111] region near the [111] any rotation decreases the distance and the probability of rotation. The region between the [211] and the [111] directions is a special case in which, although the rotations lead away from the [111] corner point, the end point of the rotations, the [211], is still sufficiently near this corner point for the probability of rotation to be comparatively low. This should lead to a spread of orientations between [111] and [211], with a tendency for concentration near each of these directions.

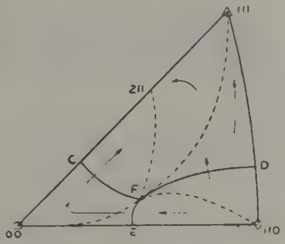
Fig. 4 shows the general trends of the rotations which give rise to the tension texture in fig. 4. There appears to be no experimental work on pure tension textures with which these orientation spreads may be compared, and it has been the custom to consider the textures developed in cold-drawn wires as tension textures. While it is reasonable to suppose that in the wire drawing process longitudinal tension may be the predominating factor, it is probable that the ideal tension texture will be modified by radial compressive stresses, and discussion of this effect will be left until after some consideration of compression textures.

Fig. 3.



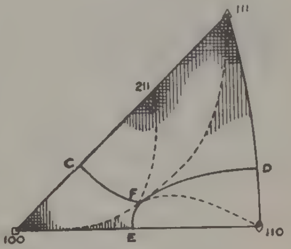
Lines of quickest descent from the shear stress peak (F), and directions of rotation due to single crystal slip (→I) and duplex slip (→II) in a face-centred cubic metal in tension.

Fig. 4.



General trends of the rotations in a face-centred cubic metal in tension.

Fig. 5.



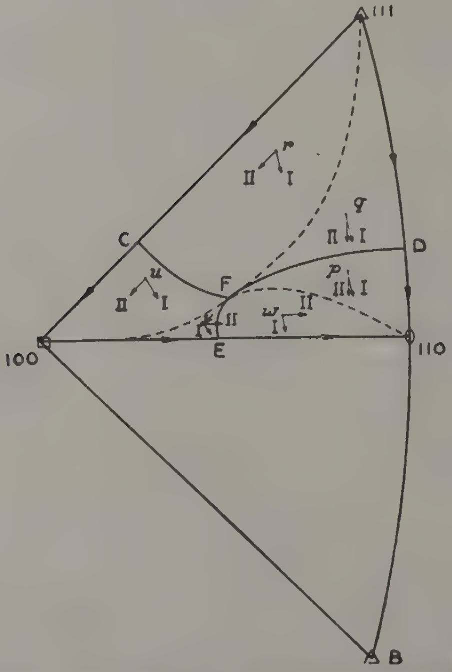
Distribution of grain orientations after tension as represented by their applied stress directions (tension texture).

COMPRESSION TEXTURE.

The compression texture may be derived in an exactly similar manner to that described above, bearing in mind the fact that the rotation directions for single and duplex slip are not simply reversed. For single slip the applied stress direction moves now towards the pole of the operative slip plane, while under duplex slip it moves towards the great circle joining the two operative poles. Thus for an orientation represented by the point *r* in fig. 3 and fig. 6 the single crystal rotation is in the tension case towards the operative [110] slip direction A, and in the compression case towards the pole of the operative (111) slip plane B. The duplex rotations are however merely reversed except in the small region between the [111] and [211] directions.

Considering now the rotations in more detail, it will be seen that the orientations contained in the area F[111][110] will move towards the line F[110] and, in addition, an appreciable proportion of the orientations in CF[111] will also move towards this line. This will clearly denude the regions suronding [111]. Some of those in the CF[111] region, namely, those near to CF, will move across this line. In the region CF[100] there is some tendency for rotation towards the [100] so that a small residual concentration of orientations may be expected near to [100]. Most of the orientations here however, together with those in the EF[100] region, will move in the EF[110] region. In this latter region the path of T_e is primarily along the line E[110], and consequently duplex rotation towards [110] is the more probable. This will also have the effect of turning those orientations which have reached the F[110] boundary from the area above the line sharply towards the [110] direction. Fig. 7 shows these general rotations which would lead to a distribution of orientations shown in fig. 6.

Fig. 6.



Lines of quickest descent from the shear stress peak (F), and directions of rotation due to single crystal slip (→ I) and duplex slip (→ II) in a face-centred cubic metal in compression.

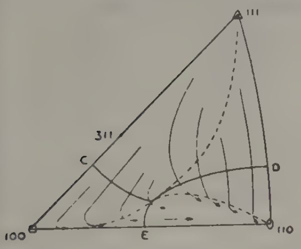
The prediction of a [110] texture is of course amply confirmed by experiments on face-centred metals and the detailed agreement between the observed distribution of orientations, fig. 8 (Barrett and Levenson 1940), and that predicted on the present treatment, is satisfactory.

Rolling Texture.

The simplest combination of stresses for the approximate simulation of the rolling process is that of a compression in the direction of the rolling plane normal together with a tension in the rolling direction (Wever 1931).

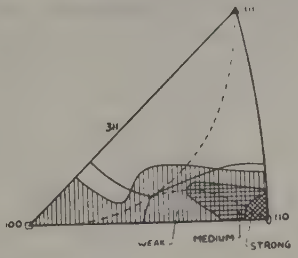
This has been adopted in the present treatment, which examines the ways in which both tension and compression end conditions may be satisfied simultaneously. The resulting rolling texture is therefore not that which would result from a mere super-position of the tension and compression textures.

Fig. 7.


111
311
C
D
100
E
110

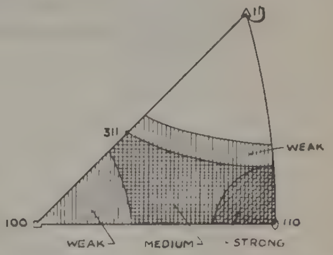
General trends of the rotations in a face-centred cubic metal in compression.

Fig. 8.


111
311
100
110
WEAK
MEDIUM
STRONG

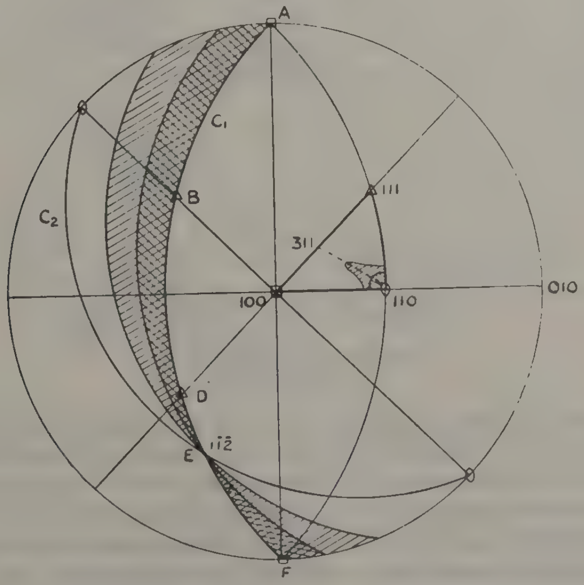
Predicted compression texture for a face-centred cubic metal.

Fig. 9.


111
311
WEAK
100
110
WEAK
MEDIUM
STRONG

Texture in aluminium after 98 per cent compression (Barrett and Levenson 1940).

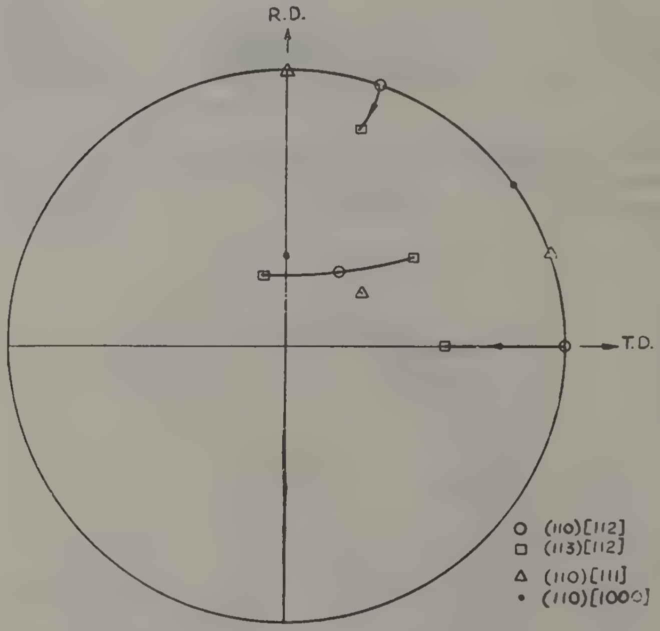
Fig. 10.


A
C₁
C₂
B
111
311
100
110
010
D
1$\bar{1}\bar{2}$
E
F

Stereographic projection showing the compression texture and the resulting spread in the rolling direction.

The compression normal to the rolling plane will produce a texture such that, for the majority of the crystals, the direction of compression will lie within the shaded area shown in the unit triangle of fig. 10. To simplify the ensuing discussion this texture may be considered as a spread of orientation along the great circle from the [110] direction towards the [311], shown dotted in the figure. Since the rolling direction is perpendicular to the rolling plane normal, the rolling direction must lie on the great circle for which the rolling plane normal is the pole. Thus the rolling direction lies on the circle C_1 for the rolling plane normal in the [110] direction, and on the circle C_2 for the rolling plane normal at the other limit, the [311] direction. The great circles corresponding to the intermediate positions of the compression spread are obtained by oscillation, about the axis $[1\bar{1}\bar{2}]$, of a great circle between the limits C_1 and C_2, and the distribution of rolling directions will be as indicated by the cross-hatching between these two great circles. At the same time, the tension in the rolling direction will tend to produce a texture similar to that illustrated in fig. 5, in which most of the crystals have the direction of tension at end-points of the form [111], [211] and [100], the tendency for a spread of orientation about these end positions being ignored at present. Such directions are shown by the points A, B, D, E and F in fig. 10. The only one of these points which completely satisfies both a tension end condition and the compression spread is E, the $[1\bar{1}\bar{2}]$ direction. Thus there will be a strong tendency for rotation to occur such that this condition is fulfilled, giving a strong (110)[112] texture. The [111] and [100] end positions of the tension

Fig. 11.

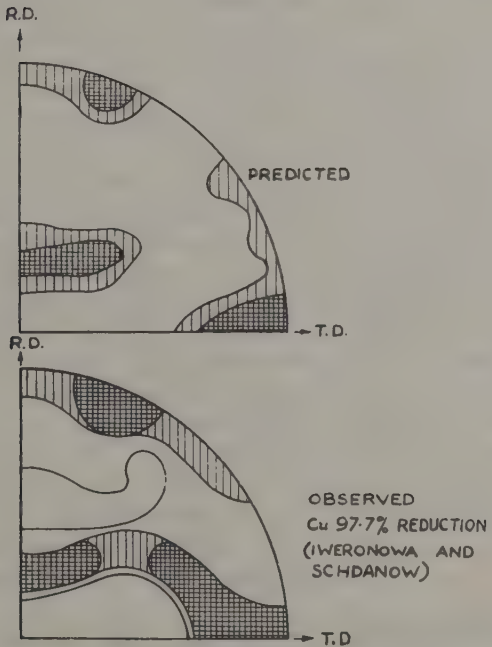


Composite (111) pole figure.

texture satisfy only the [110] end of the compression spread, but since this is the most densely populated region in the compression texture, there will be some tendency towards (110)[111] and (110)[100] components in the final rolling texture, which will be made up of the following preferred orientations.

Direction of tension (rolling direction)	Direction of compression (normal to rolling plane)	Notation
[112]	[110] → [311]	(110) [112] to (311) [112]
[111]	[110]	(110) [111]
[100]	[110]	(110) [100]

Fig. 12.

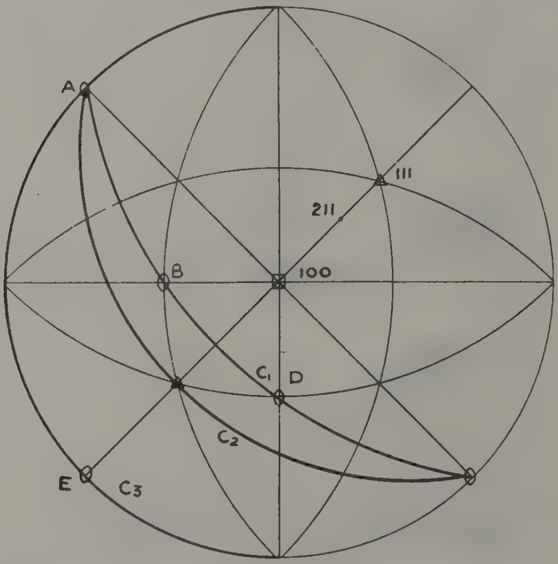


(111) pole figure showing predicted and observed rolling textures.

The ideal (111) pole figure corresponding to these four textures is shown in fig. 11. The lines joining the (111) poles of the (110)[112] and (311)[112] textures are obtained by rotation of the (110)[112] texture about the [112] direction by an amount appropriate to the compression spread from [110] to [311]. In view of the fact that only the poles corresponding to (110) [112] and (311)[112] orientations and the spread between them satisfy both the compression spread and a tension end-position these will predominate. Thus a semi-quantitative pole figure corresponding to these textures with some contributions from the (100)[111] and (110)[100] textures may be obtained as in fig. 12.

On the same figure is shown the (111) pole figure for copper after 97·7 per cent reduction, obtained by Iweronowa and Schdanow (1934). The predicted pole figure shows a much higher degree of preferred orientation, but is otherwise in good agreement with the experimental observations. It is perhaps worth noting that Barrett and Steadman (1942), in an investigation on the rolling of copper using both large grain polycrystalline material and single crystals mounted in polycrystalline blocks, found (*a*) multiple orientations developed in nearly every grain, with the exception of those with initial orientation (110)[112] which were unchanged even after 97·6 per cent reduction, and (*b*) the entire polycrystalline texture developed in a crystal with the initial orientation (100)[001]. This supports the view that the (110)[112] orientation will predominate.

Fig. 13.



Stereographic projection showing ideal tension texture and the corresponding range of compression directions.

DRAWING TEXTURE.

The drawing process will be considered as equivalent to a tension exerted on the metal in the direction of drawing together with uniformly distributed radial stresses in the plane perpendicular to the drawing direction. The longitudinal tension, considered to be the predominating influence, will tend to produce the tension texture already described, namely, concentrations of orientations near [111], [211] and [100] directions. The compression directions, being perpendicular to the tension direction, must be at 90° from these end points, and therefore lie on the great circles C_1, C_2 and C_3 respectively, fig. 13. A compression at any point on these circles attempts to bring a [110] direction, the compression texture being regarded for simplicity as predominantly at [110] texture, into alignment with itself whilst retaining the tension textures. For any point on C_1 this requires a rotation of not more than 30° to the points A, B, and D, and on C_3 of not more than 45° to A and E. On C_2, however, a rotation of as much as 90° to A would apparently be demanded, so it is far more likely that the orientations corresponding to C_2 will, except for those near to A, rotate to join the [110] end points B and D. This will reduce the density of orientations around [211] in the tension texture, and correspondingly augment the [111] texture, thus producing the double fibre texture having [111] and [100] as the fibre axes.

The experimental results of Schmid and Wassermann (1927) show the existence of this double texture in copper, silver and gold, although the proportions of the two textures vary from metal to metal. Aluminium however shows only the [111] component.

It is not possible to predict from purely geometrical considerations these variations in texture, but in the present treatment the development of textures depends on the extent to which the effective stress, T_e, can move before deformation occurs, that is to say, on the ability of the boundary regions of individual grains or sub-units of structure to withstand large stress gradients. This property may well be dependent on the temperature of the experiment in relation to the melting point of the metal and its purity ; the nearer to the melting point and the higher the purity, the less likely is the metal to withstand these stresses. If the movement of T_e is so limited that mainly single crystal rotation occurs the texture will be different from that developed in another face-centred cubic metal in which duplex rotation is the predominating feature. When mainly single crystal rotations occur the [100] component of both the compression and tension textures will be absent. Thus considering the development of the drawing texture, fig. 13, C_3 will be absent and a single fibre axis, the [111], will result. It may be significant that aluminium, which at room temperature is nearer its melting point than the other face-centred metals, copper, silver and gold, does indeed show this single texture.

CONCLUSIONS.

From this qualitative treatment of the deformation of polycrystalline agregates it has been possible to derive for face-centred cubic metals :—

(1) a tensile stress-elongation curve which is in satisfactory agreement with experiment and with that previously derived by Taylor,

(2) tension, compression, rolling, and wire-drawing textures which are in reasonable agreement with observation.

An attempt has been made to give a clearer physical basis of the deformation process and since the above results can be obtained without laborious calculations it may be possible to extend the treatment to body-centred cubic metals.

A very tentative suggestion is put forward to explain the variation of deformation textures among metals of the same crystal structure.

ACKNOWLEDGMENTS.

The authors wish to acknowledge the encouragement and helpful comments received from Dr. N. P. Allen, and the assistance given by the Mathematics Division of the Laboratory in the calculation of the Resolved Shear Stress Tables.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES.

BARRETT, C. S., and LEVENSON, L. H., 1940, *Trans. A.I.M.E.*, **127,** 112.
BARRETT, C. S., and STEADMAN, F. W., 1942, *Trans. A.I.M.E.*, **147,** 57.
BOAS, W., and SCHMID, E., 1931, 1, *Z. tech. Physik*, **12,** 71.
CALNAN, E. A., and BURNS, B. D., 1950, *J. Inst. Metals*, **77,** 445.
GÖLER, V., and SACHS, G., 1929, *Z. Physik*, **55,** 581.
IWERONOWA, W., and SCHDANOW, G., 1934, *Tech. Phys. U.S.S.R.*, **1,** 64.
MISES, R. v., 1928, *Z. angew. Math. Mech.*, **8,** 161.
PICKUS, M. R., and MATHEWSON, C. H., 1939, *J. Inst. Metals*, **64,** 237.
SCHMID, E., and WASSERMANN, G., 1927, *Z. Physik*, **42,** 779.
TAYLOR, G. I., 1938, *J. Inst. Metals*, **62,** 307 ; Stephen Timoshenko, 60th Anniversary Volume, p. 218, Macmillan, New York.
WEVER, F., 1931, *Trans. A.I.M.E.*, **93,** 51.

XCVIII. *The Energy Distribution behind Decaying Shocks.—I. Plane Waves.*

By M. J. LIGHTHILL*.

[Received July 11, 1950.]

SUMMARY.

A hypothesis is made concerning the degree of accuracy of a theory of decaying plane shocks due to Friedrichs. This hypothesis is carefully examined in the light of energy balance ; it is substantiated by the checks made on it, and in the course of the investigation the picture of the disturbance to the fluid as a whole is considerably improved. The work will be used in Part II. to clarify other phenomena involving decaying shocks, including some in steady supersonic flow.

Most of the analysis refers to the following problem. A plane wave, headed by a shock, is produced by a piston, which is moved impulsively into homogeneous stationary fluid and then retarded arbitrarily until it comes to rest. One may usefully ask what is the whole history of the disturbance to the fluid if viscosity and heat conduction be neglected, since the modifying effects of these can be sketched in afterwards. The Friedrichs theory replaces the disturbance by a " simple wave " (with the fluid behind it in its original undisturbed state) and satisfies conditions at the shock neglecting the cube of its strength (see §§2–4 below). A precise hypothesis about the degree of accuracy of this theory is made in §1, but there is no experimental or logical certainty that this is correct at large distances from the piston. A comprehensive check is made as follows, arbitrary thermodynamic properties being assumed for the fluid.

The energy of the simple wave at any time t (§6) is compared with the work done by the piston (§5). The difference (§7) is a quantity of the second order in the initial shock strength, obtained with an error which (on the accuracy hypothesis) is of the third order. This residual energy $R(t)$ can only represent energy of the fluid behind the simple wave (where disturbances are of the third order but their integrated effect may be of the second order). It is comparable with the total entropy gain $S(t)$ of the fluid behind the simple wave, which is known from the theoretical variation of shock strength. In fact the relation

$$R(t)=T\left(1-\frac{1}{2}\frac{T_V P}{TP_V}-\frac{1}{8}\frac{PP_{VV}}{P_V^2}\right)S(t) \quad . \quad . \quad . \quad . \quad (1)$$

is found to hold, neglecting third order terms, where the suffix derivatives are taken with respect to minus the volume at constant entropy (see §2 for notation).

---

* Communicated by the Author.

This rather surprising relation is checked as follows. The residual energy is divided (§ 8) into that due to changes in entropy and pressure separately, the former changes being known. Hence the integrated pressure over the fluid behind the simple wave is determined at all times. Since also these small variations of pressure satisfy the equation of sound (even though the entropy is non-uniform) and the pressure gradient at the stationary piston is zero, this determines completely the third order pressure waves in this part of the fluid. They are such as would be produced if the shock emitted a pressure pulse of magnitude

$$\frac{1}{16A}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)T\frac{dS}{dt} \quad . \; . \; . \; . \; . \; . \; . \quad (2)$$

backwards through the fluid at all times t, which was then reflected at the piston. The bracket in (2) is negative for most fluids but zero for monatomic gases.

The wave (2) is next shown (§ 9) to consist of the reflection of the simple wave at the shock. The reflection coefficient in this process may be calculated exactly by considering the shock conditions with the local equations of motion, but only the third order terms in the reflected wave can be accurate after it has interacted with the simple wave. These are precisely (2), which completes the check by showing that the conclusions of the theory in the large agree with those in the very small.

To demonstrate that the said third-order terms do not change by interaction with the simple wave, the important, but apparently unrecognized, fact is used that, in fairly weak waves, a small variation of entropy is taken account of to a first approximation if the ordinary theory, at constant entropy, is applied with the velocity and *pressure* as independent variables ; in particular the flow behind a shock can then be obtained correctly to the third order in the strength. This is explained in § 9.

The extension to an " N-wave ", produced by the piston being retarded through zero to a negative velocity, and then stopping abruptly to produce a second shock, is made in § 10. The accuracy hypothesis, which is again substantiated by the checks made, now implies, however, that the motion of the rear shock after a long time is given less accurately by the theory than that of the front shock. Until the shocks have decayed to a small fraction of their initial strength, the accuracy is unimpaired ; also equations (1) and (2) still hold. The pressure waves (2) are now due to the combined effects of reflection of the simple wave at the front shock and transmission of it through the rear shock. But the reflections of these pressure waves at the piston gradually overtake the rear shock as time goes on and changes its path from that given by the Friedrichs theory (equation (79) shows by how much). Thus, when they are rarefaction waves, they communicate their negative energy to the N-wave, by slowing down the rear shock and so widening the N-wave on its low-pressure side. After a very long time, all the residual energy will be due to the increase of total entropy of the fluid with, ultimately, no change in pressure; *i. e.* $R(\infty)=(1-T_VP/TP_V)S(\infty)$. The remaining energy is propagated to infinity in the N-wave (whose width increases as its amplitude decreases).

It is shown in §11 that, if the piston has stopped on returning to its original position, then after an intermediate length of time the ratios of the energy in the N-wave, in the increased entropy, and in the rear pressure waves, are as $5:10:(-1)$ for air. The third form of energy is subsequently transferred to the first. On the other hand, if the piston stops much ahead of the original position, the first-named form of energy always predominates.

---

§1. INTRODUCTION.

DECAYING shocks appear, for example, whenever there is an explosion. Also, when any object moves through fluid faster than sound, the bow shock is observed to become weaker as we follow its course away from and behind the object; this may also be called decay. A number of papers have been written on the mechanics of the process (see References). Some give detailed theories of its progress in particular cases. These mostly apply only when the shock is fairly weak, and hence only to the later stages of decay for a strong shock. Taylor (1950) treats the early stages of decay for a very strong explosion.

A careful study of the published theories reveals certain doubts, difficulties, and even divergences of opinion. These are connected with considerations of energy and entropy. It is the purpose of this paper, with Part II., to clarify these points, and to answer such questions as (i) When a wave, with a decaying shock in front, and possibly one behind also, passes through fluid, what relationship, if any, is there between the rate of gain of energy, and of entropy, of the fluid behind the wave? (ii) If either is substantial, can a theory of the propagation of the wave, which makes use of assumptions valid only under adiabatic conditions, have any value?

No reference to any but a single type of decaying shock will now be made until §10. The interval will be filled with a very detailed analysis, in this one representative case, of the questions broached above.

Consider a very long tube indeed, of unit cross-sectional area, filled with homogeneous fluid at rest. Let a piston at one end (say $x=0$) begin to move into it impulsively, with a certain finite velocity, at time $t=0$, and be subsequently retarded in an arbitrary manner until after a certain time it is again at rest. Then at any instant, if the tube is long enough, the fluid in motion will be confined between the piston and a plane shock. The shock will decay in strength as it moves along the tube faster than sound.

Viscous stress and heat flow by conduction will be very small, except inside the shock, and where otherwise they are largest (near the front of the wave) will not act on any element of fluid for very long. A theoretical treatment which neglects them, and replaces the very thin shock by a plane surface of discontinuity, should therefore be quite satisfactory. Then the motion is governed by Euler's equations of continuity and momentum, the Rankine-Hugoniot relations at the shock, and constancy of entropy for each element of fluid after the shock has passed it by ; the boundary condition is that the velocity of the fluid in contact with the piston is that of the piston.

No exact mathematical solution is known, even to a particular case of this problem. An approximate theory of Friedrichs * should be good, according to plausible hypotheses, when the velocity of the piston is always rather less than the velocity of sound in the undisturbed fluid, so that the shock is always weak. But there has been no experimental verification, or mathematical proof, of the validity of the approximate theory. It would not perhaps be difficult to produce either if the approximations were required to be good only within a limited distance of the piston. But there are serious difficulties, even in the mathematical discussion, at arbitrarily large distances from the piston, where any terms neglected may have had an arbitrarily long time to influence the propagation of the wave. (It may be noticed here that, the weaker be the shock, the longer does it take to decay to, say, one half of its strength.)

The Friedrichs method is to determine the " simple wave ", at the entropy of the undisturbed fluid, which is compatible with the motion of the piston. This is taken as the flow behind the shock. (In a simple wave the velocity and the thermodynamic variables are all functionally related. Any value of one of them is propagated through the fluid at a constant speed, namely at the local speed of sound relative to the local fluid velocity. The simple wave can exist only when the specific entropy is uniform. In the present problem it will terminate with the sound wave emitted by the piston as it comes to rest. Behind this, on simple wave theory, the fluid is once more homogeneous and at rest.) The shock is calculated approximately as a plane surface moving in such a way that the physical quantities in the simple wave thereat take those values which the motion of the shock into fluid at rest would produce, with the cube of the shock strength neglected. The calculations were given by Friedrichs only for the case when the fluid is a perfect gas with constant specific heats, but the extension to a fluid with general thermodynamic properties is a fairly simple matter.

It would be impossible to satisfy the conditions at the shock more exactly ; for example, the specific entropy, which has to be taken uniform in the simple wave, must actually increase at the shock by an amount of order the cube of the strength.

---

* Friedrichs, 1948, pp. 220–231. This theory of the phenomenon, which is the subject of the present §§ 1–9, will be constantly referred to as " the Friedrichs theory ", even though in the paper cited many other phenomena were also investigated.

Friedrichs (1948) is cautious about claiming any precise degree of accuracy for this theory, but an attempt at estimating the errors may be made. These errors arise from using the incorrect value of the entropy, and from the fact that only a second order boundary condition can be applied at the shock. These considerations suggest the following

Accuracy Hypothesis.—The errors in the Friedrichs theory are comparable with the changes produced by altering the simple wave in two respects : (i) the specific entropy of the whole wave is altered to its known maximum, *i. e.* by an amount of order the cube of the maximum shock strength ; (ii) the velocity distribution is altered, say by changing the velocity of the piston (continuously, and even after it has actually come to rest) by an amount whose ratio to the velocity of sound is of order the cube of the maximum shock strength.

If this hypothesis is correct, the Friedrichs theory is of considerable value at all distances from the piston, as will be seen in §4. Even after many multiples of the duration of motion of the piston, the position of the shock (relative to its position if it has been an acoustic wave) is in error only by a factor one plus a term of order the square of the maximum shock strength. Even though this factor multiplies a quantity which grows indefinitely, the approximation may be regarded as highly satisfactory.

The accuracy hypothesis might be questioned on two grounds, both based on the fact that the total entropy gain of the fluid as a whole, as a result of the passage of the shock, is of order the square of the initial shock strength *. (i) The effect on the wave of the terms neglected, if it accumulates as time goes on, might be held as a consequence to be of order greater than the cube of the maxiumm shock strength. (ii) It might be expected that the total energy of the fluid is defective (on the Friedrichs theory) by an amount of order the square of the maximum strength (and this is true), and that this indicates errors of such an order in the wave.

The author rejects objection (i) on the grounds that the form of the accuracy hypothesis already permits propagation of errors along the diverging characteristics, and hence their progressive accentuation ; thus, in the model simple wave used for comparison, the effects of the new value for the specific entropy accumulate as time goes on. The defective total energy mentioned in (ii) is the subject of this paper, which will analyse it in detail, assuming the accuracy hypothesis. It is sufficient here to remark that, though errors within the simple wave itself could not account for this deficiency in total energy, the region behind it (which the Friedrichs theory gives as undisturbed) is permitted, on the accuracy hypothesis, to include disturbances of order the cube of the maximum shock strength. The integrated effect of these over the extensive region behind the simple wave is sufficient to account for the discrepancy in energy.

---

* That this is so, although the gain in specific entropy of any element is of order the cube of the strength, is because the weaker shocks take longer to decay.

But the hypothesis remains unproved, and it was to make a further test of its correctness, and in particular to make a detailed critique of objection (ii), that the further deductions from the hypothesis, given below, were made. If they had led to a contradiction, or to physically unplausible results, the hypothesis would have had to be modified. In fact, the results are neither impossible or unlikely, and the considerably fuller picture of the flow as a whole, which this investigation in the large provides, is confirmed in quantitative detail by an investigation of the local reflection of the simple wave at the shock.

The deductions which are to be made have already been sketched in the Summary. At any time t the total increase in energy, kinetic and internal, of the fluid in the simple wave over its internal energy in the undisturbed state—briefly, " the energy in the simple wave "—will be calculated on the accuracy hypothesis. Now the energy in the whole fluid must equal the work done by the piston. Hence, if the accuracy hypothesis is true, the difference must represent energy of the fluid behind the simple wave. From the expression for this residual energy, the pressure distribution in this region is deduced, to a first approximation, as a third order function varying only relatively gradually. The same pressure distribution is then deduced in a totally different manner, as a result of reflection of the simple wave at the shock followed by reflection at the piston.

The author has thought it worth while to set out the work for a fluid with perfectly general thermodynamic properties, for three reasons:—(i) The proportionality of residual energy and entropy shown in equation (1) (in the absence of which the theory might find itself in difficulties), and the agreement in expression (2) as obtained by two methods, then appear as the result of the cancellation of a large number of terms, in a manner impossible to believe fortuitous. (ii) The work shows up explicitly what terms arising in it are due to departures from adiabatic conditions * (roughly those bringing in the temperature). (iii) It is desirable to place on record a statement of the Friedrichs theory for fluids other than perfect gases.

The " tail " behind the simple wave is found to be a compression wave, in which the slightly reduced pressure directly behind the simple wave is gradually made up to its undisturbed value. Technically this compression wave must ultimately form a shock, but it is so weak that this could not happen for a very long time, and need hardly be considered in detail †. One may also notice that the increased entropy in the general neighbourhood of the piston, which the theory leaves as permanent, would gradually be flattened out, for a real fluid, by conduction of heat, causing a further slight increase in the total entropy, though not in the total energy.

---

[*] See § 3 for a discussion of this point, and of the errors in a paper of Martin due to his ignoring it.

[†] The indication that an N-wave must ultimately form, even in the present problem, is, however, interesting.

§2. The Simple Wave.

The notation to be used for the various physical quantities, at points in the undisturbed fluid and elsewhere, is indicated in the table below.

Quantity	Undisturbed	Disturbed
Pressure	P	$P+p$
Specific volume	V	$V-v$
Velocity	0	u
Velocity of sound	$A=V\sqrt{P_V}$	$a=(V-v)\sqrt{(\partial p/\partial v)_s}$
Temperature	T	.....
Internal energy	E	$E+e$
Specific entropy	0	s

Derivatives, in states not that of the undisturbed fluid, will be indicated in the manner usual in thermodynamics; thus $(\partial p/\partial v)_s$ indicates a derivative at constant entropy. But the values in the undisturbed state of derivatives along an adiabatic (*i. e.* at constant entropy) will be denoted by suffixes and by the use of capital letters; P_V is the value of the derivative just mentioned in the undisturbed fluid, P_{VV} the corresponding second derivative, and so on *. Suffix S will also be used to denote differentiation with respect to entropy at constant volume in the undisturbed state.

The Euler equations of continuity and momentum are

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+(V-v)\frac{\partial u}{\partial x}=0, \qquad \frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+(V-v)\frac{\partial p}{\partial x}=0. \quad . \quad (3)$$

Hence at constant entropy

$$\left.\begin{aligned} \frac{a}{V-v}\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}\right)-\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}\right)+a\left(\frac{\partial u}{\partial x}-\frac{a}{V-v}\frac{\partial v}{\partial x}\right)=0,\\ \frac{a}{V-v}\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}\right)+\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}\right)+a\left(\frac{\partial u}{\partial x}+\frac{a}{V-v}\frac{\partial v}{\partial x}\right)=0. \end{aligned}\right\} \quad . \quad (4)$$

A solution is therefore given by putting

$$u=\int_0^v a\frac{dv}{V-v} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (5)$$

to satisfy the first equation, so that the second becomes

$$\frac{\partial u}{\partial t}+(u+a)\frac{\partial u}{\partial x}=0, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (6)$$

showing that u is constant along straight characteristics $dx/dt=u+a$. This solution is the " simple wave ".

---

* Notice that P_V is therefore *minus* the derivative of pressure with respect to volume (and P_{VVV} minus the third derivative, but P_{VV} plus the second derivative). This is convenient as making the quantities all positive for most fluids.

Friedrichs showed that it was convenient to regard the simple wave as specified by a single function $\sigma(\xi)$, which he defined as follows. ξ is a parameter constant on each straight characteristic, namely the value of x where the characteristic produced would meet the axis $t=0$. It is always negative in the problem of this paper. Then, writing $u+a=\mathrm{A}(1+\sigma)$, it follows, since u and a are constants on any characteristic, that σ and ξ are functionally related. Actually, in the present problem, it is more convenient to regard ξ as a function of σ than vice versa, for it is a problem in which this function is one-valued, unlike some considered by Friedrichs. Then the equation of a characteristic is

$$x=\xi(\sigma)+\mathrm{A}t(1+\sigma). \qquad \ldots\ldots \qquad (7)$$

Since σ is a non-dimensional measure of the disturbance to the fluid, its cube may always be neglected in the Friedrichs theory.

All the physical quantities are related in the simple wave ; all will now be expressed in terms of σ, neglecting cubes of disturbances. Firstly, by (5),

$$u=\int_0^v \frac{\mathrm{A}+\mathrm{A}_\mathrm{V}v}{\mathrm{V}-v}\,dv=\frac{\mathrm{A}}{\mathrm{V}}v+\frac{1}{2}\left(\frac{\mathrm{A}}{\mathrm{V}^2}+\frac{\mathrm{A}_\mathrm{V}}{\mathrm{V}}\right)v^2, \qquad \ldots\ldots \qquad (8)$$

whence

$$\sigma=\frac{u+a-\mathrm{A}}{\mathrm{A}}=\left(\frac{1}{\mathrm{V}}+\frac{\mathrm{A}_\mathrm{V}}{\mathrm{A}}\right)v+\frac{1}{2}\left(\frac{1}{\mathrm{V}^2}+\frac{\mathrm{A}_\mathrm{V}}{\mathrm{AV}}+\frac{\mathrm{A}_\mathrm{VV}}{\mathrm{A}}\right)v^2. \qquad \ldots \qquad (9)$$

Replacing A by $\mathrm{VP}^{\frac{1}{2}}_\mathrm{V}$, and its derivatives accordingly, (9) may be rewritten

$$\sigma=\frac{\mathrm{P}_\mathrm{VV}}{2\mathrm{P}_\mathrm{V}}v-\frac{1}{2}\left(\frac{\mathrm{P}_\mathrm{VV}}{2\mathrm{VP}_\mathrm{V}}-\frac{\mathrm{P}_\mathrm{VVV}}{2\mathrm{P}_\mathrm{V}}+\frac{\mathrm{P}^2_\mathrm{VV}}{4\mathrm{P}^2_\mathrm{V}}\right)v^2, \qquad \ldots\ldots \qquad (10)$$

whence
$$v=\frac{2\mathrm{P}_\mathrm{V}}{\mathrm{P}_\mathrm{VV}}\left[\sigma+\left(\frac{1}{2}+\frac{\mathrm{P}_\mathrm{V}}{\mathrm{VP}_\mathrm{VV}}-\frac{\mathrm{P}_\mathrm{V}\mathrm{P}_\mathrm{VVV}}{\mathrm{P}^2_\mathrm{VV}}\right)\sigma^2\right] \qquad \ldots \qquad (11)$$

and
$$u=\frac{2\mathrm{AP}_\mathrm{V}}{\mathrm{VP}_\mathrm{VV}}\left[\sigma+\left(1+\frac{\mathrm{P}_\mathrm{V}}{\mathrm{VP}_\mathrm{VV}}-\frac{\mathrm{P}_\mathrm{V}\mathrm{P}_\mathrm{VVV}}{\mathrm{P}^2_\mathrm{VV}}\right)\sigma^2\right]. \qquad \ldots \qquad (12)$$

The coefficient of σ^2 in u vanishes in the case considered by Friedrichs, of a perfect gas with constant specific heats.

According to the accuracy hypothesis the errors in (7), (11) and (12) are such as would be produced by adding terms of order σ^3 (or rather the maximum of this quantity) to those in σ and σ .

§ 3. CONDITIONS AT THE SHOCK.

The values of disturbed quantities behind the shock are related by equations expressing conservation of mass momentum and energy for an element of fluid as the shock passes through it, say with velocity U. These are

$$\frac{\mathrm{U}}{\mathrm{V}}=\frac{\mathrm{U}-u}{\mathrm{V}-v}, \qquad p=\frac{\mathrm{U}}{\mathrm{V}}u, \qquad (\mathrm{P}+p)u=\frac{\mathrm{U}}{\mathrm{V}}\left(\tfrac{1}{2}u^2+e\right). \qquad \ldots \qquad (13)$$

Elimination of U and u gives the Hugoniot relation between the thermodynamic variables

$$(\mathrm{P}+\tfrac{1}{2}p)v=e. \qquad (14)$$

It is convenient to expand (14) in series, using that $\mathrm{P}+p=(\partial e/\partial v)_s$. The best form is as ,

$$[\mathrm{E_V}+\tfrac{1}{8}(\mathrm{E_{VVV}}v^2+2\mathrm{E_{VVS}}vs+\mathrm{E_{VSS}}s^2)]v=\mathrm{E_V}v+\mathrm{E_S}s$$

$$+\frac{1}{24}(\mathrm{E_{VVV}}v^3+3\mathrm{E_{VVS}}v^2s+3\mathrm{E_{VSS}}vs^2+\mathrm{E_{SSS}}s^3)+\mathrm{O}(v^5+s^5), \quad . \quad (15)$$

where (in this one equation only in the whole paper) suffix derivatives refer not to the undisturbed state but to a state with specific volume and entropy halfway between those ahead and of behind the shock. (In this way quadratic and quartic terms on both sides of the equation are avoided.) This gives the entropy as

$$s=\frac{\mathrm{E_{VVV}}}{12\mathrm{E_S}}v^3=\frac{\mathrm{P_{VV}}}{12\mathrm{T}}v^3, \qquad (16)$$

the error being $\mathrm{O}(v^5)$ if $\mathrm{P_{VV}}$ and T are taken in the said intermediate state, but $\mathrm{O}(v^4)$ if (as will be sufficient in what follows) they are taken in the undisturbed state.

Equation (16) is a convenient form of the relation between the thermodynamic variables behind the shock ; it states how near to the adiabatic this relation is. The velocities u and U are deducible, by (13), as

$$u=(pv)^{\frac{1}{2}}, \qquad \mathrm{U}=\mathrm{V}(p/v)^{\frac{1}{2}}. \qquad (17)$$

It should be noticed that it follows from these equations, with (16), that in the expansion of u in powers of p or v alone, neglecting cubes of the disturbances, the adiabatic relation may be used, but that this is not so in the expansion of U, which is

$$\mathrm{U}=\mathrm{V}[\mathrm{P_V}+\tfrac{1}{2}\mathrm{P_{VV}}v+\tfrac{1}{6}\mathrm{P_{VVV}}v^2+\ \ldots\ +\mathrm{P_S}(s/v)+\ \ldots]^{\frac{1}{2}}$$

$$=\mathrm{A}\left[1+\frac{\mathrm{P_{VV}}}{2\mathrm{P_V}}v+\left(\frac{\mathrm{P_{VVV}}}{6\mathrm{P_V}}+\frac{\mathrm{T_V}}{\mathrm{P_V}}\frac{\mathrm{P_{VV}}}{12\mathrm{T}}\right)v^2\right]^{\frac{1}{2}}, \qquad (18)$$

since $\mathrm{P_S}=\mathrm{E_{VS}}=\mathrm{T_V}$.

Martin (1947) falls into the error of neglecting the effect on U of the entropy change. Hence second order terms in his calculations are wrong. One result of this mistake is that he obtains a reflection of a simple wave at a shock, whereas in fact, for the special case of a perfect gas with adiabatic index 5/3 which he considers, the reflection coefficient is at most of the fourth order, as will be seen in §9. Friedrichs avoids the mistake by using the explicit solution for a perfect gas of the Rankine-Hugoniot relation. The $\mathrm{T_V}/\mathrm{T}$ in (18), which will recur in many subsequent formulae, is a constant reminder that in one place, at least, the Friedrichs theory does have to take account of the entropy change at the shock.

§ 4. DETERMINATION OF THE SHOCK.

The shock condition must now be expressed, using the "simple wave" values (§2) for the various physical quantities behind it. Then v is related to the fundamental variable σ by (11), and the expansion (18) becomes

$$U=A(1+\tfrac{1}{2}\sigma+\lambda\sigma^2),$$

where

$$\lambda=\frac{1}{8}+\frac{P_V}{6P_{VV}}\left(\frac{3}{V}-\frac{P_{VVV}}{P_{VV}}+\frac{T_V}{T}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

For a perfect gas with constant specific heats λ is simply $\frac{1}{8}$.

At time t let the variable σ take the value σ_1 at the shock, so that σ_1 is a measure of the strength of the shock at time t. Then the condition $dx/dt=U$ along the shock, with (7) and (19), gives

$$(\xi'(\sigma_1)+At)\frac{d\sigma_1}{dt}+A(1+\sigma_1)=A(1+\tfrac{1}{2}\sigma_1+\lambda\sigma_1^2), \quad . \quad . \quad . \quad (20)$$

whence

$$(\tfrac{1}{2}\sigma_1-\lambda\sigma_1^2)A\frac{dt}{d\sigma_1}+At=-\xi'(\sigma_1). \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

The initial condition for the shock is $x=0$ at time $t=0$, and so by (7) $\xi=0$ when $t=0$. If $\sigma=\sigma_0$ initially, so that σ_0 is a measure of the maximum shock strength, then (21) may be solved exactly as

$$At=2\left(\frac{1-2\lambda\sigma_1}{\sigma_1}\right)^2\int_{\sigma_1}^{\sigma_0}\frac{\sigma\,d\xi(\sigma)}{(1-2\lambda\sigma)^3}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

but, since cubes of σ have been neglected in the formulæ on which this equation is based (without which it would be impossible to satisfy the shock conditions completely) the only significant terms in (22) are

$$At=\left(\frac{2}{\sigma_1^2}-\frac{8\lambda}{\sigma_1}\right)\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi(\sigma)+\frac{12\lambda}{\sigma_1^2}\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi(\sigma). \quad . \quad . \quad . \quad (23)$$

This tells us the time t at which the non-dimensional measure of disturbance σ takes the value σ_1 just behind the shock. It is easy to verify that the condition (17) on u is also satisfied to the required order, by comparing it with (5).

Here the significance of the errors in (23) which the accuracy hypothesis permits should be noticed. They are terms like

$$\sigma_1^{-2}\int_{\sigma_1}^{\sigma_0}\sigma^3\,d\xi, \qquad \sigma_1^{-1}\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi, \qquad \int_{\sigma_1}^{\sigma_0}\sigma\,d\xi, \text{ etc.}$$

Their ratio to the terms included is always of order σ_0^2, although the first two become indefinitely large as time goes on and σ_1 decreases towards zero. In consequence, the value of σ_1 for given t is also in error only by a factor $1+O(\sigma_0^2)$, as stated in §1.

§ 5. THE PISTON.

It is necessary to have an expression for the motion of the piston in terms of the fundamental function $\xi(\sigma)$. The former should logically come first, as being the cause of the whole motion, but it is much more convenient to work from σ throughout. It will be sufficient for our purposes to neglect the square of σ in determining the time t at which the piston emits the characteristic specified by σ.

At the piston $dx=u\,dt$; hence, by (7, and (12),

$$d\sigma(\xi'(\sigma)+\mathrm{A}t)+\mathrm{A}\,dt\left(1+\sigma\left(1-\frac{2\mathrm{P_V}}{\mathrm{VP_{VV}}}\right)\right)=0, \quad . \quad . \quad . \quad (24)$$

whose solution, still neglecting σ^2, is

$$\mathrm{A}t=-\xi(\sigma)+\sigma\xi(\sigma)+\frac{2\mathrm{P_V}}{\mathrm{VP_{VV}}}\int_{\sigma}^{\sigma_0}\sigma\,d\xi, \quad . \quad . \quad . \quad . \quad (25)$$

taking $x=0$, and so $\xi=0$, initially (when $\sigma=\sigma_0$).

Where the piston comes to rest $u=0$, and hence, by (12), $\sigma=0$ (neglecting, on the accuracy hypothesis, σ_0^3). The "simple wave" will be taken as the region satisfying $0<\sigma<\sigma_1$ exactly, *i. e.* lying between the shock and the characteristic whose speed is that of sound in the undisturbed fluid. This characteristic is not emitted precisely as the piston comes to rest, but (as will be seen in § 9) very slightly earlier.

The work done by the piston against the pressure of the fluid, neglecting σ_0^3, is

$$\int_{\sigma_0}^{0}(\mathrm{P}+p)u(dt/d\sigma)\,d\sigma$$

$$=\int_{0}^{\sigma_0}(\mathrm{P}+\mathrm{P_V}v)\frac{2\mathrm{AP_V}}{\mathrm{VP_{VV}}}\left(\sigma+\left(1+\frac{\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_V P_{VVV}}}{\mathrm{P_{VV}^2}}\right)\sigma^2\right)\frac{1}{\mathrm{A}}$$

$$\times\left(\xi'(\sigma)-\xi(\sigma)-\left(1-\frac{2\mathrm{P_V}}{\mathrm{VP_{VV}}}\right)\sigma\xi'(\sigma)\right)d\sigma$$

$$=\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left[\int_0^{\sigma_0}\sigma\,d\xi+\left(\frac{1}{2}+\frac{2\mathrm{P_V^2}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_V P_{VVV}}}{\mathrm{P_{VV}^2}}\right)\int_0^{\sigma_0}\sigma^2\,d\xi\right], \quad . \quad . \quad . \quad . \quad (26)$$

where (12) and (25) have been used and, in integrating $\sigma\xi(\sigma)$ by parts, the fact that it vanishes at both limits.

To the order of approximation employed, (26) must represent the total gain in energy of the fluid at any time after the piston has come to rest.

§ 6. ENERGY OF THE SIMPLE WAVE AT TIME *t*.

The energy of the simple wave will now be calculated, on the accuracy hypothesis, at any time t after the piston has come to rest. Under adiabatic conditions the increased kinetic and internal energy per unit mass of a disturbed element of fluid is

$$\tfrac{1}{2}u^2+\int_{\mathrm{V}}^{v}(\mathrm{P}+p)\,dv,$$

which in the simple wave is

$$\tfrac{1}{2}u^2+\mathrm{P}v+\tfrac{1}{2}\mathrm{P_V}v^2=\frac{2\mathrm{PP_V}}{\mathrm{P_{VV}}}\left[\sigma+\left(\frac{1}{2}+\frac{\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}+\frac{2\mathrm{P^2_V}}{\mathrm{PP_{VV}}}\right)\sigma^2\right], \qquad (27$$

neglecting σ^3.

The total energy in the simple wave $0<\sigma<\sigma_1$, obtained by integrating (27) multiplied by $dx/(\mathrm{V}-v)$, with x given by (7), is

$$\frac{2\mathrm{PP_V}}{\mathrm{P_{VV}}}\int_0^{\sigma_1}\left[\sigma+\left(\frac{1}{2}+\frac{\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}+\frac{2\mathrm{P^2_V}}{\mathrm{PP_{VV}}}\right)\sigma^2\right]\left(\frac{1}{\mathrm{V}}+\frac{2\mathrm{P_V}}{\mathrm{V^2P_{VV}}}\sigma\right)$$

$$\times(\xi'(\sigma)+\mathrm{A}t)\,d\sigma=\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left[\int_0^{\sigma_1}\sigma\,d\xi+\mathrm{A}t\tfrac{1}{2}\sigma_1^2+\left(\frac{1}{2}+\frac{2\mathrm{P^2_V}}{\mathrm{PP_{VV}}}\right.\right.$$

$$\left.\left.+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}\right)\left(\int_0^{\sigma_1}\sigma^2\,d\xi+\mathrm{A}t\tfrac{1}{3}\sigma_1^3\right)\right]. \quad . \quad (28)$$

With At expressed in terms of σ_1 by (23), this energy becomes

$$\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left[\int_0^{\sigma_0}\sigma\,d\xi-4\lambda\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi+6\lambda\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi\right.$$

$$\left.+\left(\frac{1}{2}+\frac{2\mathrm{P^2_V}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}\right)\left(\int_0^{\sigma_1}\sigma^2\,d\xi+\tfrac{2}{3}\sigma_1\int_0^{\sigma_1}\sigma\,d\xi\right)\right] \quad . \quad (29)$$

Since in (28) At is multiplied entirely by terms σ_1^2, σ_1^3, etc. *, the error in applying (23) to obtain (29) is at most of order σ_0^3 (arising from integrals like $\int_0^{\sigma_0}\sigma^3\,d\xi$). Thus, on the accuracy hypothesis, (29) is the correct first and second order expression for the energy of the simple wave.

§7. THE RESIDUAL ENERGY.

The difference between (26) and (29) must therefore represent energy in the region behind the simple wave. Subtracting, one obtains for this residual energy

$$\mathrm{R}(t)=\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left(\frac{1}{2}+\frac{2\mathrm{P^2_V}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P^2_{VV}}}-6\lambda\right)$$

$$\times\left(\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi-\tfrac{2}{3}\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi\right). \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Equation (30) has been proved only at times t after the piston has come to rest. But, before this, At=O(1), and study of (23) shows that then $|\xi(\sigma_1)|=\mathrm{O}(\sigma_0)$, so that (30) is $\mathrm{O}(\sigma_0^3)$. The residual energy at this early stage may also be shown to be of the third order, and so (30) remains true in the sense that the difference between its left- and right-hand sides is always a third order quantity.

---

* This is because the simple wave has been considered as terminating with the characteristic on which $\sigma=0$ exactly. This point will be examined further in § 9.

It will be observed that all integrals in (26) and (29) involving the range $0<\sigma<\sigma_1$ have disappeared by the cancellation of eight distinct terms. This suggests that $R(t)$ depends only on the properties of the *shock* previous to time t. The total entropy of the fluid at time t, say $S(t)$, also has this property, and a comparison of the two is clearly desirable.

Unit mass of fluid, on passing through the shock, gains an entropy s given by (16), and subsequently retains it. To a first approximation mass passes through the shock at a rate A/V; also v is given by (11) as $2P_V\sigma/P_{VV}$. Hence the rate of increase of total entropy of the fluid is

$$\frac{dS}{dt}=\frac{A}{V}\,\frac{P_{VV}}{12T}\left(\frac{2P_V}{P_{VV}}\right)^3\sigma_1^3. \qquad (31)$$

Hence, using the first approximation to (23) to integrate (31),

$$S(t)=\int_{\sigma_0}^{\sigma_1}\frac{dS}{dt}\,\frac{dt}{d\sigma_1}\,d\sigma_1=\frac{A}{V}\,\frac{P_{VV}}{12T}\left(\frac{2P_V}{P_{VV}}\right)^3\int_{\sigma_0}^{\sigma_1}\sigma_1^3\,\frac{1}{A}\left(-\frac{2\xi'(\sigma_1)}{\sigma_1}-\frac{4}{\sigma_1^3}\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi\right)d\sigma_1$$

$$=\frac{2}{3}\,\frac{P_V^3}{VTP_{VV}^2}\left(6\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi-4\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi\right). \qquad (32)$$

Comparison of (32) with (30) shows that $R(t)$ and $S(t)$ are proportional at all times (neglecting third order terms). Substitution for λ in (30) from (19), and division, gives their relationship in the form already stated in the summary, namely

$$R(t)=T\left(1-\frac{1}{2}\,\frac{T_VP}{TP_V}-\frac{1}{8}\,\frac{PP_{VV}}{P_V^2}\right)S(t). \qquad (1\text{ bis})$$

The " coincidences " necessary to obtain this relatively simple relationship already add to one's confidence in the Friedrichs theory. On the other hand, the relationship is not so simple that any rough-and-ready arguments, based on familiarity with the thermodynamics of systems in equilibrium, would have been able to predict it.

§8. STATE OF THE FLUID BEHIND THE SIMPLE WAVE.

According to the accuracy hypothesis, the velocities, and the disturbances to the thermodynamic variables, are of the third order in the region behind the simple wave. The total energy is $R(t)$. Since kinetic energy terms are of the sixth order, they are negligible in building up the total energy, and so

$$R(t)=\frac{1}{V}\int_0^{At}e\,dx. \qquad (33)$$

Here the limits are not exactly correct for describing the region behind the simple wave, but the error involved in using them in (33) is still of the third order, since the true boundaries are exactly parallel to $x=0$ and $x=At$ respectively.

Now, to a first approximation,

$$e = \mathrm{T}s + \mathrm{P}v = \mathrm{T}s + \mathrm{P}\left(-\frac{\mathrm{P_S}}{\mathrm{P_V}}s + \frac{1}{\mathrm{P_V}}p\right). \quad . \quad . \quad . \quad . \quad (34)$$

Hence

$$\int_0^{\mathrm{A}t} p\,dx = \frac{\mathrm{P_V}}{\mathrm{P}}\left\{\mathrm{VR}(t) - \left(\mathrm{T} - \frac{\mathrm{PT_V}}{\mathrm{P_V}}\right)\mathrm{VS}(t)\right\}$$

$$= \tfrac{1}{8}\mathrm{T}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}} - \frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\mathrm{S}(t) \quad . \quad . \quad . \quad . \quad . \quad (35)$$

by (34), (33) and (1). Thus the fluid behind the simple wave cannot be entirely at its original undisturbed pressure unless the bracket in (35) is zero (so that the bracket in (1) becomes $1 - \mathrm{T_V P/TP_V}$) *.

For a perfect gas with constant adiabatic index γ the bracket in (35) is $3\gamma - 5$. Thus it vanishes for monatomic gases, but is negative for other gases, even if they are only approximately perfect, since for these γ is always considerably less than $\frac{5}{3}$. The bracket is probably not positive for any fluid. Thus for liquids the second term, corresponding to the relative decrease of compressibility with decreasing specific volume, is always considerable, exceeding about 3 even at high pressures, and at least 6 at normal pressures. On the other hand, the first term may be written, in terms of the coefficient of expansion α and the compressibility κ, as $4\alpha\mathrm{V}/c_v\kappa$. Here $4\mathrm{V}/c_v$ is limited by the law of Dulong and Petit to less than about 3×10^{-7} cm.^{2}/dyne.deg. But α/κ, the pressure increase due to raising a rigidly enclosed gram of liquid through one degree, is not as much as twenty atmospheres at normal pressures, as (it is seen) would be needed to make the first term in the bracket attain to the lower limit 6 of the second term. On the other hand, at high pressures the coefficient of expansion falls off rapidly, while the compressibility increases, indicating that here too the bracket in (35) is negative. It will be assumed, throughout what follows, that the bracket is non-positive †, as has here been indicated may be true for all, or nearly all, fluids. Then, by (35), the average pressure behind the simple wave is below the undisturbed value (or equal to it, to the third order, for a monatomic gas, a gloss which will not be repeated henceforth in similar contexts).

Now in this region of small variations of pressure, entropy and velocity, those of *pressure* at least are governed by the equation of sound. For by (3), with small disturbances,

$$\frac{\partial v}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} = 0, \qquad \frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial p}{\partial x} = 0. \quad . \quad . \quad . \quad . \quad . \quad (36)$$

---

* Nor can the fluid be at its original undisturbed specific volume, which would mean that (1) is replaced by $\mathrm{R}(t) = \mathrm{TS}(t)$.

† In words, the conjecture is that the slope of an adiabatic, as graph of pressure against volume, increases at least as rapidly as the fourth power of the temperature.

But entropy is effectively independent of t, because constant for a given fluid element. Hence $\partial v/\partial t=(\mathrm{V}^2/\mathrm{A}^2)\partial p/\partial t$, and, therefore,

$$\frac{\partial^2 p}{\partial t^2}=\mathrm{A}^2\frac{\partial^2 p}{\partial x^2}. \qquad (37)$$

Also, by the second of equations (3), at the piston (where $x\simeq 0$, and $u=0$ at all times t after the piston has come to rest),

$$(\partial p/\partial x)_{x=0}=0. \qquad (38)$$

Equations (35), (37) and (38) determine p completely (to the present third order approximation). By (37) and (38)

$$p=f(\mathrm{A}t+x)+f(\mathrm{A}t-x) \qquad (39)$$

for some f. Equation (35) then gives

$$\int_0^{2\mathrm{A}t} f(y)\,dy=\tfrac{1}{8}\mathrm{T}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\mathrm{S}(t). \qquad (40)$$

Hence, behind the simple wave,

$$p=\frac{\mathrm{T}}{16\mathrm{A}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\left[\mathrm{S}'\left(\frac{1}{2}\left(t+\frac{x}{\mathrm{A}}\right)\right)+\mathrm{S}'\left(\frac{1}{2}\left(t-\frac{x}{\mathrm{A}}\right)\right)\right], \qquad (41)$$

where $\mathrm{S}(t)$ is given by (31) *.

The first term in square brackets in (41) represents a pressure wave propagated in the direction x decreasing, opposite to that of the shock. The second term is the reflection of the first at the piston. At $x=\mathrm{A}t$ the first is simply $\mathrm{S}'(t)$. Since the shock satisfies $x=\mathrm{A}t$ with a proportional error which is uniformly small, we may say that the pressure fluctuations behind the simple wave are such as would be produced if at each time t the shock sent backwards a pressure pulse

$$\frac{\mathrm{T}}{16\mathrm{A}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\frac{d\mathrm{S}}{dt}=\frac{\mathrm{P_V^3}}{24\mathrm{VP_{VV}^2}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\sigma_1^3, \qquad (42)$$

as stated in the Summary, which is subsequently reflected at the piston.

Since $\mathrm{S}'(t)$ decreases monotonically to zero, it follows that after a long time the pressure (41) is only considerable within a fringe or " tail ", say $\mathrm{A}(t-t_1)<x<\mathrm{A}t$. behind the simple wave, wherein $\mathrm{S}'(\frac{1}{2}(t-x/\mathrm{A}))$ is still significant. Thus only the reflected wave is of importance after a long time. The main simple wave " over-expands " slightly to a negative pressure

$$\frac{\mathrm{T}}{16\mathrm{A}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\mathrm{S}'(0), \qquad (43)$$

---

* It may be remarked that, by (36), the velocity takes a form similar to (41), but with an additional factor V/A and with the sign of the first term in square brackets changed.

for large t (and to twice this value for small t), and this pressure is recovered during the passage of a very much wider compression wave. At every point

$$\lim_{t \to 0} p = 0 ;$$

thus the initial undisturbed pressure is finally recovered. The entropy, however, remains permanently altered, unless, as pointed out at the end of §1, heat conduction is taken into account.

§9. CAUSE OF THE BACKWARD-EMITTED PRESSURE WAVE.

It has been inferred, by a somewhat roundabout investigation of the disturbance to the fluid as a whole, that the shock emits a pressure wave backwards, the pulse emitted at time t being given by (2), or (42). This will now be verified directly by a study of conditions just behind the shock. This will show that the pressure wave may be regarded as a reflection of the simple wave at the shock.

Immediately behind a plane shock, which moves into undisturbed fluid, the pressure may locally be divided into two waves each moving with the local speed of sound relative to the local fluid velocity, one in the direction of motion of the shock and the other in the opposite direction. The ratio of the pressure gradient across the latter to that across the former may be called the reflection coefficient of waves at the shock. It may be calculated exactly for shocks of arbitrary strength, and is proportional to the cube of the strength for weak shocks. It does not seem to have been given explicitly in the literature, though an essentially similar problem, of two-dimensional steady supersonic flow, was solved by the author (Lighthill 1949 a).

The reflected pressure wave is of the same order of magnitude as the entropy variations behind the shock. However, it remains unchanged to this order as it passes through the simple wave into the fluid behind. This is because *pressure and velocity*, unlike the other variables involved, *remain insensitive to variations of entropy to a first approximation*—i.e. *ignoring products of variations of entropy with those of other quantities.*

For sound waves this fact is known, and has been used in §8 ; but this section is concerned with waves of finite, though not large, amplitude. Special cases of the result have been used by the author in his work on the flow behind a stationary shock (Lighthill 1949 a) and on the diffraction of blast (Lighthill 1949 b, 1950), but he has realized its truth in the generality given above only more recently.

The proof fills only one paragraph. In the full equations of continuity and momentum for an inviscid fluid, derivatives of the pressure with respect to all three space variables occur. But the only derivative of the density which occurs is the " total " derivative following a particle of fluid. Now, neglecting conduction and radiation of heat, the entropy is constant for a given particle of fluid, so that this total derivative of density may be replaced by that of the pressure, divided by the square of the velocity of sound. The four equations are then homogenous in the first derivatives of the pressure and the velocities, the coefficients of these derivatives being the velocities and various thermodynamic variables. Hence, in replacing the latter by their values as functions of the pressure on a particular adiabatic, only variations in entropy multiplied by variations in velocity or pressure are neglected in the equations. So, if such products are negligible, it is sufficient to solve the equations of continuity and momentum for the velocity and pressure as if the entropy were constant.

Thus in the case of plane waves the Riemann theory can be used to determine the velocity and pressure. It will perhaps add to the reader's conviction if this fact is deduced independently from the exact characteristic equations for plane waves (Courant and Friedrichs 1948, p. 82):

$$\left.\begin{array}{c} du=-\dfrac{\mathrm{V}-v}{a}dp \quad \text{on} \quad \dfrac{dx}{dt}=u+a, \qquad du=+\dfrac{\mathrm{V}-v}{a}dp \quad \text{on} \quad \dfrac{dx}{dt}=u-a, \\ ds=0 \quad \text{on} \quad \dfrac{dx}{dt}=u. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (44)$$

If in these equations $(\mathrm{V}-v)/a$ be replaced by the value as a function of p which it takes for one particular value of the entropy, say $(\mathrm{V}-v)/a=\mathrm{G}'(p)$, changes in u and p will be given errors of order the entropy multiplied by those changes themselves. So if variations in u and p are first order and those of entropy are third order, the errors in taking

$$\left.\begin{array}{l} u+\mathrm{G}(p)=\text{constant} \quad \text{on} \quad \dfrac{dx}{dt}=u+a, \\ u-\mathrm{G}(p)=\text{constant} \quad \text{on} \quad \dfrac{dx}{dt}=u-a, \end{array}\right\} \quad . \quad . \quad . \quad (45)$$

(to replace the ordinary Riemann equations) will be of the fourth order. (But if $\mathrm{G}(p)$ were replaced by a function of v which is identical along the adiabatic $s=0$—in fact by the integral (5)—the error would be of the third order.)

Now consider equations (44) at a point directly behind the shock. Here the shock conditions (14) and (17) hold. Given any one of u, p, v, these determine the others. Also at adjacent points on the shock the conditions hold. Hence changes of u along the shock are related in a specific way to changes of p; that is du/dp along $dx/dt=\mathrm{U}$ is a known function deducible from conditions (14) and (17). To be precise, it is

$$\left(\frac{du}{dp}\right)_{\text{shock}}=\frac{(\mathrm{P}-(\partial e/\partial v)_p)v+(\partial e/\partial p)_v p}{2u(\mathrm{P}+\frac{1}{2}p-(\partial e/\partial v)_p)}. \quad . \quad . \quad . \quad . \quad (46)$$

But if $dp/dt=\alpha$ on $dx/dt=u+a$, so that $du/dt=-\alpha(\mathrm{V}-v)/a$, and if $dp/dt=\beta$ on $dx/dt=u-a$, so that $du/dt=\beta(\mathrm{V}-v)/a$, then

$$\left(\frac{du}{dp}\right)_{\text{shock}}=\frac{(du/dt)_{dx/dt=\mathrm{U}}}{(dp/dt)_{dx/dt=\mathrm{U}}}$$

$$=\frac{\dfrac{\mathrm{U}-u+a}{2a}\left(-\dfrac{\alpha(\mathrm{V}-v)}{a}\right)+\dfrac{-\mathrm{U}+u+a}{2a}\left(\dfrac{\beta(\mathrm{V}-v)}{a}\right)}{\dfrac{\mathrm{U}-u+a}{2a}\alpha+\dfrac{-\mathrm{U}+u+a}{2a}\beta}$$

$$=\frac{\mathrm{V}-v}{a}\,\frac{(a+u-\mathrm{U})\beta-(a-u+\mathrm{U})\alpha}{(a+u-\mathrm{U})\beta+(a-u+\mathrm{U})\alpha}. \qquad \text{(47)}$$

This gives the reflection coefficient exactly as

$$\frac{\alpha}{\beta}=\frac{a+u-\mathrm{U}}{a-u+\mathrm{U}}\left(1-\frac{a}{\mathrm{V}-v}\left(\frac{du}{dp}\right)_{\text{shock}}\right)\Big/\left(1+\frac{a}{\mathrm{V}-v}\left(\frac{du}{dp}\right)_{\text{shock}}\right) \qquad \text{(48)}$$

For a weak shock the numerators in (48) are respectively of the first and second orders in the strength, so that the whole is of the third order. For since, by (16), behind the shock

$$p=\mathrm{P_V}v+\tfrac{1}{2}\mathrm{P_{VV}}v^2+\tfrac{1}{6}\mathrm{P_{VVV}}v^3+\mathrm{P_S}\frac{\mathrm{P_{VV}}}{12\mathrm{T}}v^3+\mathrm{O}(v^4), \qquad \text{(49)}$$

while by (17) $u=(pv)^{\frac{1}{2}}$, it is easy to deduce that

$$\left(\frac{dp}{du}\right)_{\text{shock}}=\frac{\mathrm{A}}{\mathrm{V}}\left(1+\frac{\mathrm{P_{VV}}}{2\mathrm{P_V}}v+\left(\frac{\mathrm{P_{VVV}}}{4\mathrm{P_V}}+\frac{\mathrm{P_{VV}T_V}}{8\mathrm{TP_V}}-\frac{5\mathrm{P^2_{VV}}}{32\mathrm{P^2_V}}\right)v^2+\mathrm{O}(v^3)\right), \qquad \text{(50)}$$

whereas, by expanding a in powers of v,

$$\frac{a}{\mathrm{V}-v}=\frac{\mathrm{A}}{\mathrm{V}}\left(1+\frac{\mathrm{P_{VV}}}{2\mathrm{P_V}}v+\left(\frac{\mathrm{P_{VVV}}}{4\mathrm{P_V}}-\frac{\mathrm{P^2_{VV}}}{8\mathrm{P^2_V}}\right)v^2+\mathrm{O}(v^3)\right). \qquad \text{(51)}$$

Hence the reflection coefficient (48), to the third order, using (17) and (11), is

$$\frac{\alpha}{\beta}=\frac{\frac{1}{2}\mathrm{A}\sigma_1}{2\mathrm{A}}\left(\frac{\mathrm{P_{VV}T_V}}{8\mathrm{TP_V}}-\frac{\mathrm{P^2_{VV}}}{32\mathrm{P^2_V}}\right)\left(\frac{2\mathrm{P_V}}{\mathrm{P_{VV}}}\sigma_1\right)^2\Big/2=\frac{\mathrm{P_V}}{64\mathrm{VP_{VV}}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\sigma_1^3. \qquad \text{(52)}$$

The bracket in (52) is recognized as that of (2). Notice, however, that the term in $\mathrm{T_V/T}$ due to entropy change has arisen in (52) in a manner different from that in which it appeared in calculating the residual energy. It is due to the second order term in dp/du, not to that in U. This fact seems to the author to reduce still further the likelihood that the agreement of the check on Friedrichs's theory is fortuitous.

Now β, the value of dp/dt on $dx/dt=u-a$, is known approximately from the simple wave theory (on which, however, $\alpha=0$). To a first approximation $p=(2\mathrm{P^2_V}/\mathrm{P_{VV}})\sigma$; also, by (7), if $dx/dt=u-a$,

$$-\mathrm{A}\,dt=(\xi'(\sigma)+\mathrm{A}t)\,d\sigma+\mathrm{A}\,dt, \qquad \text{(53)}$$

so that

$$\beta=\frac{dp}{dt}=(2\mathrm{P^2_V}/\mathrm{P_{VV}})\frac{-2\mathrm{A}}{\xi'(\sigma)+\mathrm{A}t}. \qquad \text{(54)}$$

Hence by (52), and since for $\sigma=\sigma_1$, to a first approximation, by (20), the denominator of (54) is $-\frac{1}{2}\sigma_1\, dt/d\sigma_1$,

$$\alpha=\frac{\mathrm{P}_{\mathrm{V}}^3}{8\mathrm{VP}_{\mathrm{VV}}^2}\left(4\frac{\mathrm{VT}_{\mathrm{V}}}{\mathrm{T}}-\frac{\mathrm{VP}_{\mathrm{VV}}}{\mathrm{P}_{\mathrm{V}}}\right)\sigma_1^2\frac{d\sigma_1}{dt}. \qquad (55)$$

To see that this value of $(dp/dt)_{dx\, dt=u+a}$ is propagated through the simple wave unchanged to a first approximation, consider two adjacent characteristics $dx/dt=u-a$. The time interval Δt between them remains unchanged to a first approximation. At the shock the pressure difference between them, along a characteristic of the other family, is $\alpha\Delta t$, and the velocity difference is $-\mathrm{G}'(p)\alpha\Delta t$ by (45). As p varies on each characteristic, (45) shows that u varies by the change in $\mathrm{G}(p)$, so behind the simple wave (for example)

$$\Delta u-(-\mathrm{G}'(p)\alpha\Delta t)=\mathrm{G}'(p)(\Delta p-\alpha\Delta t). \qquad (56)$$

But if the differences between the two characteristics are again taken along a characteristic of the other family, then $\Delta u=-\mathrm{G}'(p)\Delta p$. Hence, by (56), $\Delta p=\alpha\Delta t$ as before; that is, behind the simple wave, dp/dt on a characteristic $dx/dt=u+a$ is still given by (55).

Since the shock is, to a first approximation, a characteristic of the family $dx/dt=u+a$, it has been shown that the flow behind the simple wave is the same as if the shock sent out a pulse at time t whose time derivative is (55) and which can therefore differ by at most a constant from expression (42). (Strictly it has only been shown that the waves travelling towards the piston are of this form; but waves travelling in the opposite direction must represent the reflection of the former waves at the piston if the boundary condition there is to be satisfied.)

This checks (42) to within a constant. The constant if present would, by the theory of §8, make the total energy of the fluid tend to infinity with time. But we can also see that it must be absent from the point of view of the present section. For a complete third order solution of the equations of motion and boundary conditions could now be given by specifying the pressure and velocity fields behind the shock as the combination of the simple wave, determined (to the third order) by the motion of the piston, and its third order reflection in the shock (with *its* reflection at the piston). Now the value of the pressure at the piston *immediately* after it first moves may be calculated independently, (i) from the shock conditions, and (ii) from the simple wave theory. The difference must be twice the pressure in the reflected wave at this initial stage (twice because the reflected wave is itself reflected at the piston). The two quantities are in fact the integrals with respect to u, from 0 to u, of (50) and (51) respectively. Their difference is easily verified to be twice (42). This check of (42) at $t=0$, with the previous check of its time derivative, completes the investigation of the accuracy hypothesis concerning the Friedrichs theory. It has been found watertight at every point.

The complete third order solution, which could be obtained in the manner just described, would add a negative pressure distribution to that given by simple wave theory. Now, at large distances from the piston, the reflection of the simple wave at the shock becomes negligible (by (42)), and only the reflection of this reflected wave at the piston modifies the original simple wave. This, travelling in the same direction, turns it into another simple wave (the entropy being effectively uniform at these large distances from the piston), with reduced pressure distribution. This means that the characteristics are slowed down and, therefore, that which now travels with velocity A must be one emitted slightly before the piston comes to rest. This is the one which was taken as the boundary of the simple wave in §5, for the purpose of calculating its energy in §6.

It is now clear that the form of the accuracy hypothesis adopted in §1 was necessary as well as sufficient. For a modification of the simple wave of the kind described really takes place at large distances from the piston, while near it the entropy is modified as stated. It is not surprising that the errors in the Friedrichs theory should be comparable with those obtained by introducing both these modifications, and no others, uniformly.

§ 10. THE DECAYING N-WAVE.

After the careful examination of this single class of problems, each involving only one decaying shock, it is important to discover whether the conclusions for phenomena involving more than one shock are substantially identical or not.

The problem which will be investigated is like that treated in §§1–9, except that the piston is continuously retarded (from its initial impulsively attained positive velocity) through zero speed to a negative velocity (so that it is moving away from the fluid) and is then instantaneously stopped (*not* necessarily in its original position), and subsequently held at rest. The piston, on being instantaneously stopped, emits a second shock, which moves along the tube behind the first while both gradually decay in strength. The motion of the front shock is exactly as in the problem previously studied.

Friedrichs (1948) only treats a special case, but his method is quite general. The flow between the shocks is taken to be the simple wave, at the entropy of the undisturbed fluid, which is compatible with the motion of the piston. The fluid behind the second shock is in its original undisturbed state, and the shock conditions thereat, as at the first shock, are to be satisfied to the second order in the strength. The pressure, after rising discontinuously at the front shock, falls in the simple wave monotonically to below its undisturbed value, to which it then rises discontinuously at the rear shock. This behaviour may be represented graphically as something like an N, whence the wave is called an N-wave.

Considerations as in §1 lead to the formulation of an accuracy hypothesis for the Friedrichs theory of the N-wave, with identical wording to that used in § 1. Conclusions as in §§ 2–9 can then be drawn, and will be displayed in brief, which check the hypothesis and improve the picture of the motion.

But it must be remarked at once that this hypothesis, which is about to be checked, does not forecast that the motion of the rear shock will be given to a similar degree of accuracy by the simple wave theory, at distances from the piston so large that it has decayed to a small fraction of its initial strength, as was indicated for the front shock by the theory of § 4. For the front shock, conditions on one side are known exactly. As a consequence, conditions on the other side can be specified, as in § 4, with an error of order the cube of the *local* strength. But matters at the rear shock are different. The conditions behind it are known, on the accuracy hypothesis, with a possible error of order the cube of the *maximum* shock strength. This error communicates itself to the conditions ahead of it, which determine the position of the rear shock in the simple wave; and when the shock has become very weak its motion may therefore be given quite inaccurately by simple wave theory. It will be seen later, however, by use of the improved physical picture, how to find more closely what will happen, and hence, for the time being, the consequences merely of the Friedrichs theory of the N-wave will be set out.

Taking the state behind the rear shock as undisturbed, it is easy to show that the shock conditions in the simple wave thereat have exactly the same form as for the front shock. For example, (19) holds (but notice that σ is negative at the second shock). At time t let the variable σ take the value σ_2 at the second shock. Then the differential equation (21) holds equally with σ_2 for σ_1. But the initial condition is different. If $\sigma=\sigma_3$ at the piston when it comes to rest, say at $t=t_3 \simeq -\xi(\sigma_3)/\mathrm{A}$, then these are the initial values for the second shock, at which therefore (instead of the solution (22))

$$\mathrm{A}t=2\left(\frac{1-2\lambda\sigma_2}{\sigma_2}\right)^2\left[\int_{\sigma_2}^{\sigma_3}\frac{\sigma\, d\xi(\sigma)}{(1-2\lambda\sigma)^3}+\frac{1}{2}\left(\frac{\sigma_3}{1-2\lambda\sigma_3}\right)^2\mathrm{A}t_3\right], \quad . \quad (57)$$

or, rejecting unreliable terms, as in (23),

$$\mathrm{A}t=-\left(\frac{2}{\sigma_2^2}-\frac{8\lambda}{\sigma_2}\right)\int_{\sigma_2}^{\sigma_3}\sigma\, d\xi(\sigma)-\frac{2}{\sigma_2^2}\left(6\lambda\int_{\sigma_2}^{\sigma_3}\sigma^2\, d\xi(\sigma)+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)\right). \quad (58)$$

(But since the expression (19) for U, and hence also the coefficient of $dt/d\sigma_2$ in the differential equation, may contain errors $\mathrm{O}(\sigma_0^3)$, the equation (58) is not reliable for indefinitely small σ_2.)

The work done by the piston before coming to rest is the same as in (26), but with σ_3 replacing the limit of integration 0, and hence with an additional term $\frac{1}{2}\sigma_3^2\xi(\sigma_3)$ inside the square brackets, since in integrating $\sigma\xi$ by parts the integrated term does not vanish at the limit σ_3.

At time t the simple wave is now specified by the inequality $\sigma_2<\sigma<\sigma_1$, *i.e.* it lies between the two shocks. Hence in the expression (28) for its energy the range of integration must be replaced by this one; and the integrated terms $At\frac{1}{2}\sigma_1^2$ and $At\frac{1}{3}\sigma_1^3$ must be replaced by $At(\frac{1}{2}\sigma_1^2-\frac{1}{2}\sigma_2^2)$ and $At(\frac{1}{3}\sigma_1^3-\frac{1}{3}\sigma_2^3)$ respectively. Then, replacing At by its expression (23) or (58), according as it is multiplied by powers of σ_1 or σ_2, the energy in the simple wave becomes

$$\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\Bigg[\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi-4\lambda\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi+6\lambda\int_{\sigma_1}^{\sigma_0}\sigma^2\,d\xi-4\lambda\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi+6\lambda\int_{\sigma_3}^{\sigma_2}\sigma^2\,d\xi$$
$$+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)+\left(\frac{1}{2}+\frac{2\mathrm{P_V^2}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P_{VV}^2}}\right)$$
$$\times\left(\int_{\sigma_2}^{\sigma_1}\sigma^2\,d\xi+\tfrac{2}{3}\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi+\tfrac{2}{3}\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi\right)\Bigg], \qquad (59)$$

and the residual energy, obtained by subtracting (59) from the work done by the piston, becomes

$$\mathrm{R}(t)=\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left(\frac{1}{2}+\frac{2\mathrm{P_V^2}}{\mathrm{PP_{VV}}}+\frac{3\mathrm{P_V}}{\mathrm{VP_{VV}}}-\frac{\mathrm{P_VP_{VVV}}}{\mathrm{P_{VV}^2}}-6\lambda\right)$$
$$\times\left[\left(\int_{\sigma_1}^{\sigma_0}+\int_{\sigma_3}^{\sigma_2}\right)\sigma^2\,d\xi-\tfrac{2}{3}\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi-\tfrac{2}{3}\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi\right]. \qquad (60)$$

(Again, (59) and (60) may be inaccurate, since (58) has been used to obtain them, when σ_2 has become too small.)

On the other hand the rate of increase of the total entropy $\mathrm{S}(t)$ of the fluid is given by the term (31), due to the gain at the front shock, plus a similar term, obtained by writing σ_2 for σ_1 and changing the sign, due to the gain at the rear shock; the latter is also positive since $\sigma_2<0$. Integrating with respect to the time t from the initial value $\mathrm{S}(0)=0$, and using the first approximation to (23) or (58) to express dt in terms of $d\sigma_1$ or $d\sigma_2$, the expression

$$\mathrm{S}(t)=\frac{2}{3}\frac{\mathrm{P_V^2}}{\mathrm{VTP_{VV}^2}}\left[6\left(\int_{\sigma_1}^{\sigma_0}+\int_{\sigma_3}^{\sigma_2}\right)\sigma^2\,d\xi-4\sigma_1\int_{\sigma_1}^{\sigma_0}\sigma\,d\xi-4\sigma_2\int_{\sigma_3}^{\sigma_2}\sigma\,d\xi\right], \qquad (61)$$

results. Equation (61) is correct to the second order, on the accuracy hypothesis, at all times, since after σ_2 has decreased to a small fraction of its initial value, changes in $\mathrm{S}(t)$ due to the rear shock cannot be significant.

Comparison of (61) with (60) shows that when both are true the equation (1) relating the entropy $\mathrm{S}(t)$ and the residual energy $\mathrm{R}(t)$ still holds for an N-wave. For this period of time the theory of §8, which is based simply on (1) and the accuracy hypothesis, therefore holds without change for the N-wave. Conditions behind the rear shock are, to the third order, as if at time t a pressure pulse (2) was emitted, in the direction x decreasing, from approximately $x=\mathrm{A}t$, *i.e.* (to within the degree of approximation permitted) from either of the two shocks. In terms of σ, this pulse (2) becomes

$$\frac{\mathrm{P_V^3}}{24\mathrm{VP_{VV}^2}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\sigma_1^3-\frac{\mathrm{P_V^3}}{24\mathrm{VP_{VV}^2}}\left(4\frac{\mathrm{VT_V}}{\mathrm{T}}-\frac{\mathrm{VP_{VV}}}{\mathrm{P_V}}\right)\sigma_2^3. \qquad (62)$$

(Compare (42).)

The first term in (62) corresponds to the reflection of the simple wave at the front shock, as worked out in §9. This reflection is transmitted through the rear shock together with an additional wave (corresponding to the second term in (62)) which may be regarded as the transmission of the simple wave through the rear shock. (The characteristics in the simple wave are gradually caught up by the shock, which, as it does so, touches off a small part of their energy as a transmitted wave travelling in the opposite direction.)

The analysis of the rear shock justifying the last paragraph is as follows. Ahead of it, by §9,

$$\frac{dp}{dt}=\alpha, \quad \text{and} \quad \frac{du}{dt}=-\frac{\mathrm{V}-v}{a}\alpha, \quad \text{along} \quad \frac{dx}{dt}=u+a. \quad . \quad (63)$$

But, by (54), immediately ahead of the rear shock,

$$\frac{dp}{dt}=\beta_2=(2\mathrm{P}_{\mathrm{V}}^2/\mathrm{P}_{\mathrm{VV}})\frac{4\mathrm{A}}{\sigma_2\, dt/d\sigma_2}, \quad \text{and} \quad \frac{du}{dt}=\frac{\mathrm{V}-v}{a}\beta_2, \quad \text{along} \quad \frac{dx}{dt}=u-a. \quad . \quad . \quad . \quad (64)$$

Behind the rear shock quantities are of the third order and only their first approximation is of interest; this is given by the theory of sound, on which the shock is a characteristic and, along it, say

$$\frac{dp}{dt}=\delta, \qquad \frac{du}{dt}=-\frac{\mathrm{V}}{\mathrm{A}}\delta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

Now the pressure *difference* and velocity *difference* across the rear shock are so related that changes in them satisfy (50), since the coefficients in (50) represent the state behind the shock correctly to the second order. Taken with (65) this implies that, ahead of the rear shock,

$$\left[\left(\frac{dp}{dt}\right)_{dx/dt=\mathrm{U}}-\delta\right]\Big/\left[\left(\frac{du}{dt}\right)_{dx/dt=\mathrm{U}}+\frac{\mathrm{V}}{\mathrm{A}}\delta\right]=\left(\frac{dp}{du}\right)_{\text{shock}}, \quad . \quad (66)$$

with third order terms neglected on the right and fourth order terms neglected in both numerator and denominator on the left. Substituting in the left-hand side by use of the information in (63) and (64), and maintaining the same degree of approximation, it becomes

$$\begin{aligned}\left[\frac{\mathrm{U}-u+a}{2a}\alpha+\frac{-\mathrm{U}+u+a}{2a}\beta_2-\delta\right]\Big/\Bigg[\frac{\mathrm{U}-u+a}{2a}&\left(-\frac{\alpha(\mathrm{V}-v)}{a}\right)\\ +\frac{-\mathrm{U}+u+a}{2a}&\left(\frac{\beta_2(\mathrm{V}-v)}{a}\right)+\frac{\mathrm{V}}{\mathrm{A}}\delta\Bigg]\\ =\frac{a}{\mathrm{V}-v}&\left(\alpha-\delta+\frac{a+u-\mathrm{U}}{2\mathrm{A}}\beta_2\right)\Big/\left(\frac{a+u-\mathrm{U}}{2\mathrm{A}}\beta_2-\alpha+\delta\right), \quad . \quad (67)\end{aligned}$$

so that

$$\begin{aligned}\delta&=\alpha-\beta_2\frac{a+u-\mathrm{U}}{4\mathrm{A}}\left(1-\frac{a}{\mathrm{V}-v}\left(\frac{du}{dp}\right)_{\text{shock}}\right)\\ &=\frac{\mathrm{P}_{\mathrm{V}}^3}{8\mathrm{VP}_{\mathrm{VV}}^2}\left(4\frac{\mathrm{VT}_{\mathrm{V}}}{\mathrm{T}}-\frac{\mathrm{VP}_{\mathrm{VV}}}{\mathrm{P}_{\mathrm{V}}}\right)\left(\sigma_1^2\frac{d\sigma_1}{dt}-\sigma_2^2\frac{d\sigma_2}{dt}\right), \quad . \quad . \quad . \quad (68)\end{aligned}$$

by (55), (64), (19), (51) and (50).

Integrating (68), it may be deduced that the pressure distribution along the rear side of the rear shock is as in (62). These pressures (along this approximate characteristic) may be taken as propagated acoustically along characteristics of the other family, *i.e.* towards the piston—as has been deduced previously by energy considerations alone.

It may now be seen what are the third order disturbances behind the rear shock which make its equation of motion (58) unreliable after it has greatly weakened. By this time the actual pressure pulses (62) have themselves become weak; it is the *reflection at the piston* of pulses (62) emitted at a much earlier time which are significant. (This reflected wave is given by (41) without the first term in square brackets.) Notice that ultimately, though very gradually, the whole reflected wave catches up the rear shock. For, by (58), the distance of the rear shock behind an imaginary wave emitted by the piston with velocity A on coming to rest, which is practically $-At\sigma_2$, tends to infinity like a small multiple of $t^{\frac{1}{2}}$. So all acoustic waves travelling behind it ultimately, though slowly, catch it up. Their reflection from the shock is negligible, since the reflection coefficient is of the third order. They cannot be transmitted through it, since the time decreases on both characteristics ahead of the shock. Their energy goes into an alteration of the shock path. Thus, since they are rarefaction waves, their incidence on the shock makes it weaken more slowly, and makes the N-wave become wider (on the low pressure side) than on the theory which neglects them. It is only in this indirect way that they transmit their negative energy to the N-wave.

§11. THE ENERGY DISTRIBUTION AFTER A LONG TIME.

This section will treat of the approximate energy distribution, as given by the Friedrichs theory, after the shocks have weakened to a small fraction of their initial strengths, and also of the subsequent modifications due to the gradual overtaking of the rear shock by the pressure wave behind it.

The energy distribution on the Friedrichs theory, as $t \to \infty$, will be divided into that remaining permanently in the N-wave (whose width increases as its amplitude decreases), that due to increase in entropy of the fluid through which the wave has passed (with, ultimately, no change in pressure), and that in the pressure wave behind the rear shock. It is only after these have become almost stationary at their limiting values that the gradual process of transfer of energy from the third form to the first becomes important. For this reason the said limiting values will first be determined.

The expressions which have been obtained for all these energies with t arbitrary are simplified by letting $t \to \infty$. For then, by (23) and (58), σ_1 and σ_2 tend to 0 (naturally, because the shock strengths tend to 0). The energy (59) in the simple wave becomes

$$\frac{2\mathrm{PP_V}}{\mathrm{VP_{VV}}}\left[\int_{\sigma_3}^{\sigma_0} \sigma\, d\xi + 6\lambda \int_{\sigma_3}^{\sigma_0} \sigma^2\, d\xi + \tfrac{1}{2}\sigma_3^2 \xi(\sigma_3)\right]. \quad . \quad . \quad . \quad (69)$$

The energy due to increased entropy, after the original pressure has been recovered, by (34) and (61), is

$$T\left(1-\frac{T_VP}{TP_V}\right)S(\infty)=\frac{4P_V^3}{VP_{VV}^2}\left(1-\frac{T_VP}{TP_V}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi, \quad . \quad . \quad . \quad (70)$$

and the remainder of the residual energy, corresponding to the pressure wave behind the rear shock, is

$$T\left(\frac{1}{2}\frac{T_VP}{TP_V}-\frac{1}{8}\frac{PP_{VV}}{P_V^2}\right)S(\infty)=\frac{P_V^2P}{2V^2P_{VV}^2}\left(4\frac{VT_V}{T}-\frac{VP_{VV}}{P_V}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi \quad . \quad (71)$$

by (1). For the problem of §§1–9, without a rear shock, (69), (70) and (71) are correct at all sufficiently large times. For that of §10 the gradual transfer of the energy (71) to the N-wave will make the energy of the N-wave, after a very long time indeed, equal to the sum of (69) and (71), namely to

$$\frac{2PP_V}{VP_{VV}}\left[\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi+\left(\tfrac{1}{2}+\frac{P_V}{P_{VV}}\left(\frac{3}{V}-\frac{P_{VVV}}{P_V}+2\frac{T_V}{T}\right)\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)\right]. \quad . \quad . \quad (72)$$

The energy (70) is positive for all fluids; the energy (71) is probably negative for all fluids (§8), though zero to the third order for monatomic gases. The energy (69) may be positive or negative, and normally the first term in the square brackets therein will be of a higher order of magnitude than the others, and will dominate this part of the energy and make it more important than the residual energy (70) and (71). We have then an N-wave with one " upright " of the N always taller than the other, and the internal energy on that taller side of the wave due to greater volume changes on that side predominates in importance at all times.

But there is one problem, of special interest, in which the first term in (69) does not dominate the others. This is the case when the piston is stopped precisely on reaching its initial position. It is interesting as being directly analogous to the problem of the steady two-dimensional supersonic flow past an aerofoil (neglecting boundary layer effects). The condition is that over the path of the piston

$$\int_{\sigma_0}^{\sigma_3}u(dt/d\sigma)\,d\sigma=0,$$

whence, by formulæ of §5,

$$\int_{\sigma_3}^{\sigma_0}\sigma\,d\xi+\left(\frac{1}{2}+\frac{3P_V}{VP_{VV}}-\frac{P_VP_{VVV}}{P_{VV}^2}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi+\tfrac{1}{2}\sigma_3^2\xi(\sigma_3)=0. \quad . \quad . \quad (73)$$

Then the energy (69) in the N-wave *, on substituting for λ from (19), becomes

$$\frac{2PP_V}{VP_{VV}}\left(\frac{1}{4}+\frac{P_VT_V}{P_{VV}T}\right)\int_{\sigma_3}^{\sigma_0}\sigma^2\,d\xi. \quad . \quad . \quad . \quad . \quad . \quad (74)$$

---

* This energy is still due to an inequality in the " uprights " of the N, though a smaller one than when the first term in (68) dominates the others.

Thus, if the piston comes to rest in its initial position, the approximate distribution of energy, when the shocks have decayed to a small fraction of their initial strengths, as between the N-wave, the increased entropy of the fluid as a whole, and the pressure waves behind the N-wave, is in the ratios (by (74), (70) and (71))

$$\left(\frac{1}{2}\frac{\mathrm{T_V P}}{\mathrm{TP_V}}+\frac{1}{8}\frac{\mathrm{PP_{VV}}}{\mathrm{P_V^2}}\right):\left(1-\frac{\mathrm{T_V P}}{\mathrm{TP_V}}\right):\left(\frac{1}{2}\frac{\mathrm{T_V P}}{\mathrm{TP_V}}-\frac{1}{8}\frac{\mathrm{PP_{VV}}}{\mathrm{P_V^2}}\right). \quad . \qquad (75)$$

These depend only on the physical properties of the fluid.

For a perfect gas with constant adiabatic index these ratios become

$$\frac{5\gamma-3}{8\gamma}:\frac{1}{\gamma}:\frac{3\gamma-5}{8\gamma}, \quad . \; . \; . \; . \; . \; . \; . \; . \qquad (76)$$

and for air they are

$$\frac{5}{14}:\frac{10}{14}:-\frac{1}{14}{}^{*}.$$

Notice that in this one case the total energy of the fluid is $\mathrm{TS}(\infty)$. Thus, after $\mathrm{S}(t)$ has reached its limiting value, the energy is that associated with the total entropy change when volume changes are small and integrate out to zero. That the integrated specific volume changes should be zero is inevitable, to a first approximation, since the space occupied by the fluid is unchanged if the piston returns to its initial position. Notice however that the effect is achieved by cancellation of the decreased volume in the N-wave with the increased volume (due to increased entropy without change of pressure) near the piston. Observe also that, before $\mathrm{S}(t)\simeq\mathrm{S}(\infty)$, the volume changes in the N-wave are too large, compared with the entropy changes, to be treated linearly (and in fact the energy exceeds $\mathrm{TS}(t)$, being $\mathrm{TS}(\infty)$ at all times after the piston has come to rest).

As time goes on the energy in the third form is transferred to that in the first. Ultimately (one may suppose) in all cases, and not merely approximately, all the energy which has not gone into increasing the entropy of the fluid without change of pressure is in progress to infinity with the N-wave, after that portion which has been released and reflected at the piston has once more overtaken it. (It seems unlikely that, even theoretically and after vast times, this plane N-wave would be damped by viscosity or heat conduction. For it follows from (7) that $d\sigma/dx \sim (\mathrm{A}t)^{-1}$ as $t\to\infty$, so that the gradients of all physical quantities have an approximately uniform value in the N-wave, decreasing as the reciprocal of the time. The energy dissipation in all time might then be expected to be bounded by a negligible integral of the form $\mathrm{O}\left(\int^{\infty}\mu t^{-2}\,.\,t^{\frac{1}{2}}\,dt\right)$, where μ is the viscosity).

---

* For a liquid at ordinary pressures they are practically 0 : 1 : 0, owing to the low value of $\kappa\mathrm{P}=\mathrm{P}/\mathrm{VP_V}$. (But if the piston does not return to its original position, the energy in the N-wave still predominates.)

The path of the rear shock, as it gradually moves behind the path given by the Friedrichs theory, thus increasing the width of the N-wave on the low pressure side to permit the negative energy (71) to pass into the N-wave, would be calculable to some extent, if necessary. The errors in (18), as a formula for giving the conditions at the rear shock, if the flow is disturbed behind it, are due, to a first approximation, to taking A incorrectly, and to taking U and v not relative to the velocity and volume behind the shock. Thus, if suffix 1 refers to conditions behind the rear shock, the error in (18) is $u_1+a_1-\mathrm{A}(1+\mathrm{P}_{\mathrm{VV}}v_1/2\mathrm{P}_{\mathrm{V}})$. When the shocks have weakened considerably, so that s is practically zero behind them, this expression becomes $u_1+\mathrm{V}p_1/\mathrm{A}$ approximately, a quantity constant along characteristics $dx=\mathrm{A}\,dt$, and equal by (41) to

$$\frac{\mathrm{VT}}{8\mathrm{A}^2}\left(4\frac{\mathrm{VT}_{\mathrm{V}}}{\mathrm{T}}-\frac{\mathrm{VP}_{\mathrm{VV}}}{\mathrm{P}_{\mathrm{V}}}\right)\mathrm{S}'\left(\frac{1}{2}\left(t-\frac{x}{\mathrm{A}}\right)\right). \quad . \quad . \quad . \quad . \qquad (77)$$

Adding this to (19), we see that the differential equation (21) (with σ_2 for σ_1) holds with (77) subtracted from the coefficient of $\mathrm{A}\,dt/d\sigma_2$. This equation, in which $t-x/\mathrm{A}$ is practically $-t\sigma_2$, could be treated approximately, but this will not be attempted here since the physical nature of the process is already clear, and the purposes of the paper have all been accomplished.

Some idea of the time scale of the final energy transfer is however desirable. The pulse (62) will still be considerable when σ_1 or σ_2 is, say, half its initial value, so that by (23) or (58) t is of order $\max|\xi|/\mathrm{A}\sigma_0 \simeq t_3\,\sigma_0$. Its reflection in the piston catches up with the rear shock when, at it, $\frac{1}{2}(t-x/\mathrm{A})$, *i. e.* $-\frac{1}{2}t\sigma_2$, is of this order, so that by (58)

$$\frac{t_3\sigma_0}{\sigma_2}\approx\frac{t_3}{\sigma_0}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (78)$$

Thus, when the transfer of energy from the rear pressure wave to the N-wave is in full swing, the rear shock strength is of order the square of its initial value, and the time t is of order the duration of motion of the piston t_3, divided by the cube of the initial shock strength.

The ultimate asymptotic behaviour of σ_2 (which measures the rear shock strength, and also its position, at time t) may also be obtained without difficulty. After a long time, when disturbances have become very small, only the $\mathrm{P}v$ term in the specific energy affects the integrated energy in the N-wave, and this is therefore $(\mathrm{PP}_{\mathrm{V}}/\mathrm{VP}_{\mathrm{VV}})(\sigma_1^2-\sigma_2^2)\mathrm{A}t$. Equating this expression—with $\sigma_1^2\,\mathrm{A}t$ replaced by its limiting form

$$\int_0^{\sigma_0}(2\sigma+12\lambda\sigma^2)\,d\xi$$

as given by (23)—to expression (72) gives that (neglecting σ_0^3)

$$\lim_{t\to\infty}(\mathrm{A}t\sigma_2^2)=-2\int_{\sigma_3}^{0}\sigma\,d\xi-12\lambda\int_{\sigma_3}^{0}\sigma^2\,d\xi-\sigma_3^2\xi(\sigma_3)+\left(\frac{1}{2}-\frac{2\mathrm{P}_{\mathrm{V}}\mathrm{T}_{\mathrm{V}}}{\mathrm{P}_{\mathrm{VV}}\mathrm{T}}\right)\int_{\sigma_3}^{0}\sigma^2\,d\xi. \quad . \quad . \quad . \qquad (79)$$

The Friedrichs theory gives the limit as (79) without its last term. Thus that theory describes the final stage of decay of the rear shock correctly only to a first approximation, although it describes that of the front shock correctly to a second approximation.

REFERENCES.

BRINKLEY, S. R., and KIRKWOOD, J. G., 1947, *Phys. Rev.*, **71,** 606–611.
BRODERICK, J. B., 1949, *Q.J.M.A.M.*, **2,** 98–120.
CASSEN, B., and STANTON, J., 1947, *Bull. Amer. Phys. Soc.*, **22**.
COURANT, R., and FRIEDRICHS, K. O., 1948, *Supersonic Flow and Shock Waves* (New York : Interscience Publishers, Inc.).
DUMOND, J. W. M., COHEN, E. R., PANOVSKY, W. K. H., and DEEDS, E., 1946, *J. Acoust. Soc. Amer.*, **18,** 97–118.
FRIEDRICHS, K. O., 1948, *Comm. App. Maths.*, **1,** 211–245.
LIGHTHILL, M. J., 1944, *A.R.C. R. & M.* 1930 ; 1948, *Q.J.M.A.M.*, **1,** 76–89 ; 1949 a, *Phil. Mag.*, **40,** 214–220 ; 1949 b, *Proc. Roy. Soc.* A, **198,** 454–470 ; 1950, *Ibid.*, **200,** 554–565.
MARTIN, M. H., 1947, *Q. App. Math.*, **4**, 330–348.
TAYLOR, G. I., 1950, *Proc. Roy. Soc.* A, **201,** 159–186.
WHITHAM, G. B., 1950, *Proc. Roy. Soc.* A, **201,** 89–109 ; *The Propagation of Spherical Blast* (unpublished).

XCIX. *Energy Losses of Ionizing Particles at Relativistic Velocities in the Photographic Plate.*

By H. Messel and D. M. Ritson,
Dublin Institute for Advanced Studies*.

[Received June, 26, 1950.]

The theory of ionization losses at relativistic velocities in condensed media was first treated quantitatively by Fermi (1940). Fermi showed the importance of considering the dielectric properties of the medium, and his calculations were later extended by Halpern and Hall (1948) to cover a rather general model for the dielectric properties. This theory is based on well established laws of molecular physics and classical electrodynamics, and the presence of any large deviations from it would be most surprising. In recent investigations Halpern and Hall (1948), Bradt, Kaplan and Peters (1950,) Occhialini (1950), with the photographic plate on the grain density of relativistic tracks, no increases of grain density greater than 10 per cent have been observed up to momenta of the order of 10^4 Mc. These observations have been interpreted as due either to a failure of the Halpern–Hall theory or to the hypothesis that ionization caused by distant collisions will not contribute to the development of Silver Bromide grains Occhialini (1950), Powell (1950), Wilson (1950). Both of these assumptions are difficult to accept.

However, detailed examination of the problem shows that these interpretations are based on an incorrect application of the results given by Halpern and Hall. Halpern and Hall were interested in the energy losses of particles traversing large thicknesses of matter and therefore calculated the total loss of energy in all collision processes.

The quantity required for calculations on the photographic plate as also for certain applications in cloud chamber measurements is the " probable ionization ". The development of a Silver Bromide grain will be a function of the energy actually converted into ionization losses in the grain, and will not be a function of the total energy loss. δ-rays with ranges very much greater than the radius of the grains will lose only a small fraction of their energy in the grain of Silver Bromide, and thus the energy losses, in the form of such δ-rays, of an ionizing particle will not contribute to the development of the Silver Bromide grain. The quantity required is the energy lost in all collision processes involving energies less than some maximum energy value $E_{\max}$. This quantity is defined as the " probable ionization ".

The Bethe–Bloch formula for " probable ionization " losses is as follows :

$$-\frac{dW}{dx}=\frac{2\pi ne^4}{mv^2}\left\{\log\frac{mv^2E_{\max}}{I^2Z^2(1-\beta^2)}+(1-\beta^2)\right\},$$

where all the symbols have their usual meaning.

---

* Communicated by Professor L. Jánossy.

For large values of the momentum the theory of Halpern and Hall shows the Bethe–Bloch theory overestimates the energy loss by an amount Δ given by

$$\Delta=\frac{2\pi ne^4}{mc^2}\left\{\log\frac{ne^2h^2}{\pi m\mathrm{I}^2\mathrm{Z}^2(1-\beta^2)}-1\right\},$$

thus giving an asymptotic form for the " probable " energy loss for high momenta of

$$-\frac{d\mathrm{W}}{dx}=\frac{2\pi ne^4}{mc^2}\left\{\log\frac{\pi m^2c^2\mathrm{E}_{\max}}{ne^2h^2}+1\right\}.$$

This result is identical with that given by Fermi (1940) and is quite independent of any of the assumptions as to the molecular structure of the medium.

For low momenta in the absence of conduction electrons the Halpern and Hall correction is small and thus the ratio of the " probable ionization " losses is insensitive to the minimum value for the "probable ionization" losses to any assumptions made about the molecular structure.

On detailed examination of the Halpern and Hall theory we find that the relativistic increase is entirely associated with the term related to "Cerenkov Radiation ". We have investigated the possibility that energy loss in the form of Cerenkov Radiation will be able to escape from the crystals. This contribution is given by the second term in the formula (28) of the Halpern and Hall paper. The term is

$$\frac{2\pi e^2}{v^2}\int_0^\infty -\theta v\left(\frac{1}{\epsilon}+\frac{1}{\epsilon^*}-2\beta^2\right)dv,$$

θ is a function which is 0 for $(1-\beta^2\epsilon)>0$ and π for $(1-\beta^2\epsilon)<0$. On a multifrequency theory for a dielectric ϵ will be large in the neighbourhood of a resonance frequency and less than unity or small elsewhere. Thus the contributions to the Cerenkov Radiation are mainly from frequencies in the neighbourhood of resonance lines. There will be some loss of " Cerenkov Radiation " particularly for frequencies associated with the K and L levels for which the mass absorption coefficient will be less than the order of 2000 g./cm.^{2} necessary for stopping the radiation from leaving the crystals of Silver Bromide. The theory is thus only approximate.

Fig. 1 shows a graph of the ratio of " probable ionization " to minimum " probable " ionization. For comparison purposes the dotted line shows the ratio of total ionization to minimum total ionization. The difference is due entirely to the contribution from high energy δ-rays. $\mathrm{E}_{\max}$ was taken as 5 kV. corresponding to a range of 0·2μ in Silver Bromide. We used the results of Halpern and Hall, calculated for iron. This is a fair approximation as we were concerned only with the ratios of the ionization. We note that the total relativistic increase is only 15 per cent. The reason for the difference between this result and that actually observed in a cloud chamber gas is due quite simply to the fact that the onset of the polarization effects occur at much lower momenta due to the dielectric properties of the medium. In air we should expect the polarization effects to only assume importance at momenta of the order of one hundred times greater than those for solid Silver Bromide. Thus we should expect an increase of probable ionization logarithmically up to the order of 1·7, and beyond that the ionization should become constant.

Fig. 1.

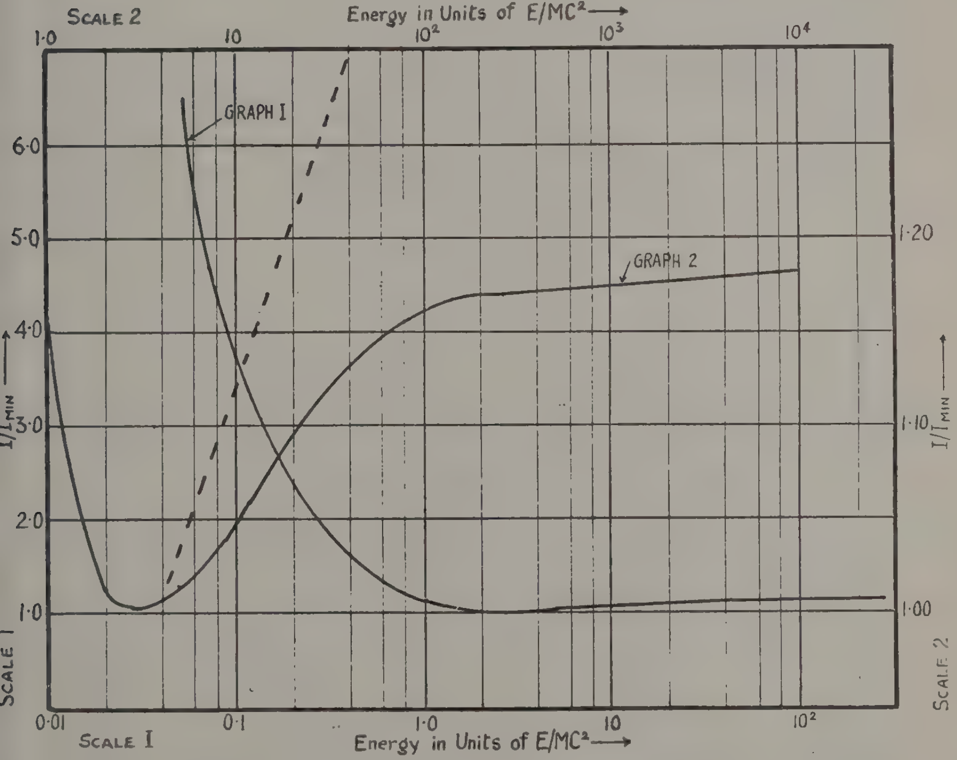


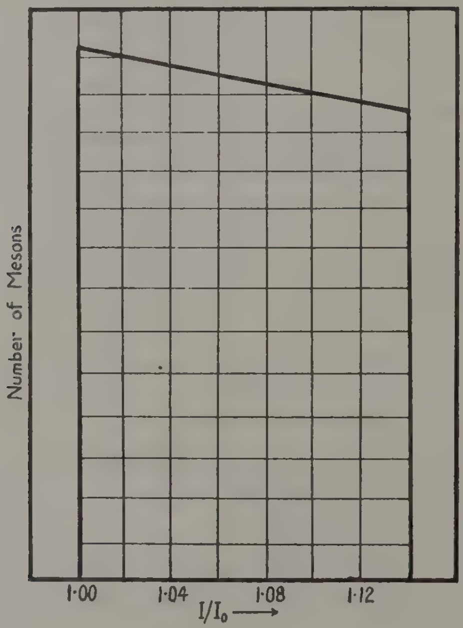
Fig. 2 shows the histogram of the ionization versus numbers calculated for the sea-level spectrum of Cosmic Rays given by Wilson (1946). We note that it is a broad distribution from 1 to 1·15 I_{min} and the mid point is at 1·07 I_{min}. The usual method of determining minimum ionization is to plot such a histogram and determine the mid point, and take this as corresponding to minimum ionization (Brown, Camerini, Fowler, Muirhead, Powell and Ritson 1949). Such a method, as can be seen by reference to fig. 11, will however determine a value of "minimum ionization" of the order of 1.07 I_{min}. The results for high altitudes are not likely to be very different. The energy spectra will contain more low energy mesons near minimum ionization, but this will be counterbalanced by more electrons with ionizations near 1·15 I_{min}.

With measurements as made at present the maximum possible rise in ionization thus comes out at about 7 per cent above the experimentally determined minimum.

Present available data gives no decisive evidence either for or against the above relativistic increase. It would however be important if it could be established with certainty. It would then provide discrimination between mesons and protons in the 10^9eV. range in measurements with the photographic plate.

We are indebted to Professor L. Jánossy and Dr. J. G. Wilson for discussions on this problem.

Fig. 2.



References

Bradt, Kaplan and Peters, 1950, *Helv. Phys. Acta*, **23**, 24.

Brown, Camerini, Fowler, Muirhead, Powell and Ritson, 1949, *Nature*, **163**, 47.

Ccchialini, G. P. S., 1950, *Bulletin du Cectre de Physique Nucléaire Bruxelles*, No. 11.

Fermi, E. 1940, *Phys. Rev.*, **57**, 485.

Halpern, O, and Hill, H, 1948, *Phys Rev.*, **73**, 477.

Powell, C. F., 1950, Edinburgh Conference, *Nature*, **165**, 54.

Wilson, J. G., 1946, *Nature*, **158**, 144.

Wilson, J. C., 1950, Edinburgh Conference, *Nature*, **165**, 54.

C. *A Theory of the Plastic Bulging of a Metal Diaphragm by Lateral Pressure.*

By R. Hill, M.A., Ph.D.,
Metal Flow Research Laboratory, Sheffield*.

[Received July 27, 1950.]

Summary.

Explicit formulæ are obtained for the stresses in a metal diaphragm which is bulged plastically by lateral pressure. The predicted influence of work-hardening on the shape of the profile, and on the relation between polar strain and curvature, agrees well with experimental data. A simple expression is developed for the instability strain.

§1. Introduction.

The ductility of sheet metal under biaxial stress is often examined by means of the so-called "bulge test". A uniform plane sheet is placed over a die with a circular or elliptical aperture and is firmly clamped around the perimeter. An increasing hydrostatic pressure is applied to one side of the diaphragm, causing it to bulge through the aperture. From the measured profile and thickness of the plastically-deformed diaphragm near the pole it is possible to calculate the local state of stress in terms of the applied pressure. If, in addition, the state of strain is measured by means of a scribed grid, the stress–strain characteristics of the metal under biaxial tension are obtained. An especial advantage of this test over any other simple one is that a greater range of pre-instability strain can be attained.

To understand the significance of this test it is desirable to have a theory which explains how the varying shape of the bulge, and the distribution of strain, depend on the properties of the metal. The theory should also be able to predict the moment when the process becomes unstable. The only published work of importance appears to be that of Gleyzal (1948) whose analysis is, however, restricted to small strains. Gleyzal formulated the necessary differential equations for a circular aperture and integrated them numerically in a particular case, but did not obtain a general solution. Moreover, he adopted physically unacceptable stress–strain relations in which the components of total strain are made to depend uniquely on the current stress. The present investigation is an attempt to obtain an explicit solution, valid for large deformations and based on the Lévy–Mises stress-strain equations.

---

* Communicated by the Author.

§2. Basic Equations.

Consider a circular diaphragm of radius a and thickness t_0, where t_0/a is so small that bending and shearing stresses can be disregarded. Let h be the polar height of the bulge when the applied pressure is p. The bulge being a surface of revolution, we can take cylindrical coordinates (r, z) for a particle, referred to axes in and perpendicular to the original plane (fig. 1). The initial radius to a particle is denoted by s, the current local thickness by t, and the inclination of the surface to the plane of the edge by θ.

By well-known geometrical formulæ the principal radii of curvature are

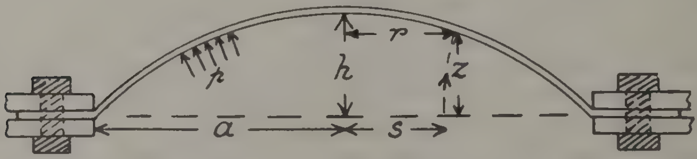
$$\rho_c = r \operatorname{cosec} \theta, \qquad \rho_t = \sec \theta \frac{dr}{d\theta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where ρ_t refers to a section such as fig. 1, and ρ_c to a perpendicular section containing the surface normal. The radii are not independent since the elimination of θ from (1) leads to

$$\frac{r}{\rho_c} \frac{d\rho_c}{dr} = 1 - \frac{\rho_c}{\rho_t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The radii are equal at the pole, but generally nowhere else unless the bulge happens to be a spherical cap.

Fig. 1.



Let σ_c and σ_t be the principal components of stress in the surface (the suffix c denotes the circumferential or hoop direction, and the suffix t denotes the tangential direction in the plane of fig. 1). The equations of equilibrium for a thin membrane in directions normal and tangential to the surface are respectively

$$\frac{p}{t} = \frac{\sigma_c}{\rho_c} + \frac{\sigma_t}{\rho_t}, \qquad \frac{\partial}{\partial r}(rt\sigma_t) = t\sigma_c. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Alternatively, these may be expressed by means of (1) and (2) as

$$\left.\begin{aligned} \sigma_c &= \frac{p}{2t} \frac{\partial}{\partial r}(r\rho_c) = \frac{p}{2t}\left(2 - \frac{\rho_c}{\rho_t}\right)\rho_c, \\ \sigma_t &= \frac{p}{2t}\rho_c \qquad\quad\; = \frac{p}{2t} r \operatorname{cosec} \theta. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The formula for σ_t may also be obtained by considering the equilibrium of any cap of the bulge. The stresses are expressed by (4) directly in terms of the pressure and the bulge geometry. Notice that if $\rho_c \gtrless \rho_t$ then $\sigma_c \lessgtr \sigma_t$.

The yield criterion is taken to be

$$f(\sigma_c, \sigma_t)=\sigma(\epsilon), \quad \ldots \ldots \ldots \quad (5)$$

where $\sigma(\epsilon)$ is the stress–strain relation in uniaxial tension or compression, ϵ is the equivalent logarithmic strain (to be defined later), and f is a homogeneous function of the stresses which does not vary with the amount of deformation.

We neglect elastic strains and adopt the Lévy–Mises relations

$$\frac{\dot{\epsilon}_c}{\dot{\epsilon}_t}=\frac{2\sigma_c-\sigma_t}{2\sigma_t-\sigma_c}, \quad \ldots \ldots \ldots \quad (6)$$

where $\dot{\epsilon}_c$ and $\dot{\epsilon}_t$ are the principal strain-rates in the surface. The condition for incompressibility is

$$\dot{\epsilon}_t+\dot{\epsilon}_c+\frac{\dot{t}}{t}=0. \quad \ldots \ldots \ldots \quad (7)$$

If $u=r-s$ is the radial displacement of a particle, and $v=\dot{u}$ is its radial speed, then

$$\dot{\epsilon}_c=\frac{v}{r}, \quad \dot{\epsilon}_t=\frac{\partial v}{\partial r}+\dot{\theta}\tan\theta. \quad \ldots \ldots \ldots \quad (8)$$

For convenience, rates of change are measured with respect to the height h, which is a suitable monotonically varying quantity. The derivative following a particle is then

$$\frac{\partial}{\partial h}+v\frac{\partial}{dr},$$

so that, for example,

$$v=\left(\frac{\partial}{\partial h}+v\frac{\partial}{\partial r}\right)u \quad \text{or} \quad v=\frac{\partial u}{\partial h}\Big/\left(1-\frac{\partial u}{dr}\right).$$

The boundary conditions on the clamped edge are

$$u, v, z=0 \text{ when } r=a, \text{ for all } h. \quad \ldots \ldots \quad (9)$$

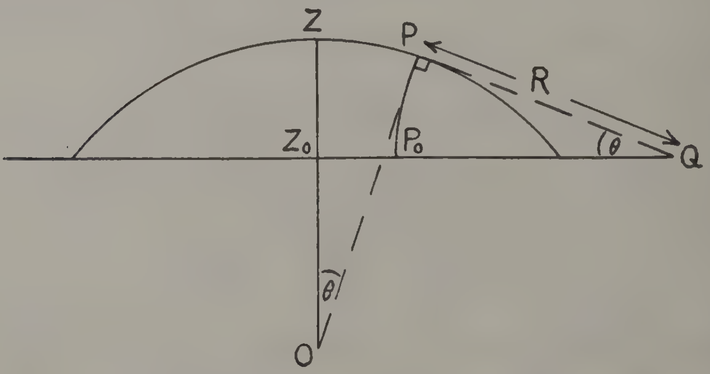
In practice the die is provided with a rounded edge to avoid premature fracture there, and the sheet is clamped around a circle of radius somewhat greater than a. The assumed boundary condition is therefore not strictly satisfied, but, when the radius of the rounded die is less than about $a/10$, the discrepancy between theory and experiment due to this cause appears to be negligible.

The condition (9) implies that the hoop strain vanishes at the edge. We shall see later that the tangential strain may or may not be zero at the edge, depending on the particular material. When the metal is thoroughly annealed, with a vanishingly small yield-stress, it is evident that the edge must strain and work-harden in order that a sufficient tangential stress is available to support the thrust of the pressure over the whole bulge. Hence, according to (6), $\sigma_c=\frac{1}{2}\sigma_t$ at the edge. The edge may, however, cease to deform later when it has become sufficiently hardened in relation to the rest of the diaphragm. On the other hand, when the metal is cold-worked and has a sharp yield-point, σ_c is not necessarily equal to $\frac{1}{2}\sigma_t$ at the edge, and deformation there is then impossible (elastic strains being disregarded). In particular, when plastic bulging is about to begin, both principal stresses are everywhere equal to the initial yield stress. The absence of peripheral strain has been observed in hard copper by Brown and Sachs (1948).

§3. A SPECIAL SOLUTION.

Before trying the general problem it is worthwhile examining whether there exists a particularly simple solution for some special stress–strain curve. Let us see what follows from the trial hypothesis that the particles describe circular paths which are, moreover, orthogonal to the momentary profile (fig. 2). Let R be the radius of the path of the particle initially at radius s. The *conventional* circumferential and tangential strains are

Fig. 2.



$$e_c=(1-\cos\theta)\frac{R}{s}, \quad e_t=-(1-\cos\theta)\left(1+\frac{dR}{ds}\right).$$

Assuming further that $e_c=e_t$, or

$$\frac{dR}{ds}+\frac{R}{s}+1=0,$$

we find that

$$R=\tfrac{1}{2}\left(\frac{a^2}{s}-s\right). \quad \text{. . . . . (10)}$$

since R=0 at the edge for zero hoop strain. This equation, being equivalent to $s(2R+s)=a^2$, is the condition that the particle paths are arcs of coaxal circles having the edges as limiting points. The successive profiles of the diaphragm are therefore the orthogonal set of circular arcs passing through the limiting points. It follows that the diaphragm is deformed into a spherical cap of radius

$$\rho=\tfrac{1}{2}\left(\frac{a^2}{h}+h\right). \quad \text{. . . . . (11)}$$

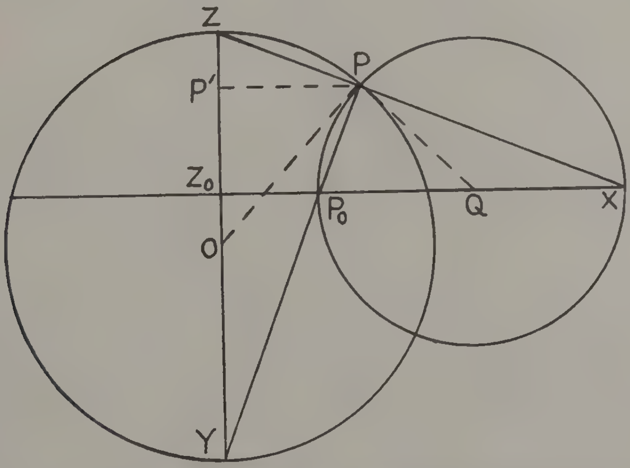
Since the principal strains in the surface are equal, the distortion of every element of the sheet is a uniaxial compression normal to the surface. The equivalent logarithmic strain ϵ is therefore $\ln(t_0/t)$. Moreover, whatever the stress–strain relations, and provided only that the material is isotropic, $\sigma_c=\sigma_t$. That is, the state of stress in every element is a balanced biaxial tension. This is equivalent to a uniaxial compressive stress if the hydrostatic component of stress is without influence on yielding. Hence (5) must reduce to

$$\sigma_c=\sigma_t=\sigma(\epsilon), \text{ where } \epsilon=\ln(t_0/t).$$

From the first equation of (3), $p=2\sigma t/\rho$ and hence the product has the same value throughout the diaphragm at each stage. This demands that the stress–strain curve is

$$\sigma=\sigma_0 \exp \epsilon, \qquad (12)$$

Fig. 3.



where σ_0 is the initial yield stress. The second equation of (3) is now satisfied identically, and the formula for the pressure becomes

$$p=2\sigma_0 t_0/\rho=4\sigma_0 t_0 h/(a^2+h^2). \qquad (13)$$

The pressure reaches its maximum when the bulge is a hemisphere. Our initial hypotheses have thus been proved consistent and we have obtained the solution for a special material. Its stress–strain curve is, however, unlike that of any actual metal in that the rate of hardening rises with increasing strain. In fact, the material extends under a constant load. None the less, the relation between the polar strain and curvature is in good agreement with that measured for various metals, as will now be seen.

We first obtain an expression for the strain at any point in terms of the polar height. Consider any two circles with centres O and Q (fig. 3) and any line through one of their points of intersection P meeting the circles in Z and X. Let ZOY and XQP_0 be the respective diameters through Z and X. Then the points P, P_0, and Y are collinear since angles ZPY and XPP_0 are both right-angles. Let Z_0 be the meet of the diameters XP_0 and ZY. Then, if the circles cut orthogonally, so do these diameters. For angle XZ_0Y is the sum of angles OZP and QXP, and therefore of angles OPZ and QPX; the sum of the latter is a right-angle since the radii OP and QP cut orthogonally, by hypothesis. Translating these results in terms of the geometry of the diaphragm, we see that if Z is the pole of the bulge and P any particle (Z_0 and P_0 being their initial positions), then PP_0 produced passes through Y, the other end of the diameter through Z. To find the final position of a particle initially at P_0, therefore, it is only necessary to produce YP_0 to cut the profile of the diaphragm.

If P′ is the foot of the perpendicular from P to OZ, the conventional hoop strain is

$$\frac{PP'}{Z_0P_0}-1=\frac{YP'}{YZ_0}-1=\frac{Z_0P'}{YZ_0},$$

by similar triangles. Hence, since $Z_0P'=z$ and $YZ_0=a^2/h$,

$$e_c=e_t=\frac{hz}{a^2}. \qquad \qquad (14)$$

The relation between the curvature and the through-thickness logarithmic strain at the pole can therefore be expressed, parametrically in terms of the polar height, as

$$\epsilon=2\ln\left(1+\frac{h^2}{a^2}\right), \quad \frac{\rho}{a}=\left(1+\frac{h^2}{a^2}\right)\Big/\frac{2h}{a}. \qquad (15)$$

Brown and Thompson (1949) found that the relations between polar curvature and strain for annealed stainless steel, annealed low-carbon steel, and certain aluminium alloys (both hard and soft), differ only slightly. The relation (15) is in excellent agreement with their mean curve up to a strain of order 0·4. It is only at greater strains that a discrepancy occurs, due partly to the occurrence of an instability and partly to the increasing dissimilarity of the stress–strain curve (12) from any actual one. The reason for the insensitivity of the $(\rho/a, \epsilon)$ relation to variations in the properties of the metal will become clear when the general solution is considered.

§4. THE INSTABILITY STRAIN.

Brown and Sachs found that when the pressure reaches its maximum a secondary bulge begins to form over about the top half of the diaphragm, and that this continues under a falling pressure. In this sense the pressure maximum can be regarded as the moment when the process becomes unstable. We can calculate the corresponding instability strain by means of (15), which we have seen to be a good approximation for a wide variety of metals. The relation between the pressure and the polar curvature is

$$p=2\sigma t/\rho \quad \text{where} \quad \epsilon=\ln(t_0/t),$$

since the local stress-state is a balanced biaxial tension equal to the yield stress σ. The pressure is a maximum when

$$\frac{1}{\sigma}\frac{d\sigma}{d\epsilon}=1+\frac{1}{\rho}\frac{d\rho}{d\epsilon}, \quad \text{. . . . .} \quad (16)$$

or, from (15), when

$$\frac{1}{\sigma}\frac{d\sigma}{d\epsilon}=\frac{3}{2}-\frac{\rho}{2h}.$$

On expanding in powers of ϵ, this reduces to

$$\frac{1}{\sigma}\frac{d\sigma}{d\epsilon}\simeq\frac{11}{8}-\frac{1}{2\epsilon}, \quad \text{. . . . .} \quad (17)$$

with an error of less than 1 per cent in the relevant range. If the stress–strain curve can be fitted by a power law $\sigma=A\epsilon^n$, the polar strain at maximum pressure is

$$\epsilon_m=\frac{4}{11}(2n+1). \quad \text{. . . . .} \quad (18)$$

For greatest accuracy the exponent n should be chosen to give best fit to the actual stress–strain curve in the neighbourhood of the instability strain.

In uniaxial tension the strain at maximum load is equal to n; in a thin spherical shell expanded by internal pressure, to $2n/3$; and in a plane sheet under balanced biaxial tension, to $2n$. In all these tests the instability strain is vanishingly small for a heavily cold-worked metal, but is equal to 4/11 in the bulge test. This strikingly demonstrates the superiority of the latter as a means of obtaining stress–strain characteristics.

Lankford and Saibel (1947) have attempted to calculate the instability strain in the bulge test by assuming (without previous analysis) that the bulge is spherical and that the strain is given by (14), but with logarithmic strain instead of conventional strain. This therefore agrees with the present analysis only when the deformation is small, which is certainly not so at maximum pressure. Lankford and Saibel's final result is still further vitiated by a gross error committed in taking partial derivatives.

§ 5. GENERAL SOLUTION FOR COLD-WORKED METALS.

The stress–strain curve of a metal in a fairly severely cold-worked state can be approximated by a linear relation

$$\sigma/Y=1+H\epsilon, \quad \text{. . . . .} \quad (19)$$

where Y is the initial yield stress and H is a non-dimensional constant of order unity. We look for a solution giving the various unknown quantities as power series in h/a. We shall take the solution only as far as the second approximation, and it will become apparent that the yield criterion is then required to be satisfied only in first approximation.

Any yield criterion (5), which reduces to von Mises' criterion for stress-states near to a balanced biaxial tension, can then be written as

$$(\sigma_c+\sigma_t)/2\mathrm{Y}\simeq 1+\mathrm{H}\epsilon, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

where the equivalent strain ϵ is equal to $\ln(t_0/t)$ to first order.

Elastic strains being neglected, bulging does not begin until the whole sheet is stressed to the yield point. Initially, therefore, $\sigma_c+\sigma_t=2\mathrm{Y}$ everywhere. The only possible stress components satisfying this condition and the second of equations (3) (with $t=t_0$) are

$$\sigma_c=\sigma_t=\mathrm{Y}.$$

This is the first approximation to the stresses. Substitution in the second of equations (4) gives $\theta=pr/2t_0\mathrm{Y}$ to first order. Now

$$\frac{dz}{dr}=-\tan\theta.$$

Hence, after substituting for θ and integrating

$$\frac{z}{h}=\left(1-\frac{r^2}{a^2}\right), \quad p=\frac{4ht_0\mathrm{Y}}{a^2},$$

on introducing the necessary conditions that $z=h$ when $r=0$, and $z=0$ when $r=a$. From (6) and (8) after substituting

$$\theta=\frac{2hr}{a^2},$$

we find that (to first approximation)

$$v=\frac{2hr}{a^2}\left(1-\frac{r^2}{a^2}\right),$$

since $v=0$ when $r=a$. Finally, from (7),

$$\ln\frac{t_0}{t}=\frac{2h^2}{a^2}\left(1-\frac{r^2}{a^2}\right).$$

This completes the solution to first approximation. To this order the bulge is spherical (or rather a paraboloid), irrespective of the rate of work-hardening.

The same sequence of solution is used for the second approximation. The yield criterion (20) becomes

$$\frac{\sigma_c+\sigma_t}{2\mathrm{Y}}=1+\frac{2\mathrm{H}h^2}{a^2}\left(1-\frac{r^2}{a^2}\right).$$

Combining this with the second of (3) we find that the second approximation to the stresses is

$$\left.\begin{aligned}\frac{\sigma_t}{\mathrm{Y}}&=1+\frac{h^2}{a^2}\left[2\mathrm{H}-(\mathrm{H}+1)\frac{r^2}{a^2}\right],\\ \frac{\sigma_c}{\mathrm{Y}}&=1+\frac{h^2}{a^2}\left[2\mathrm{H}-(3\mathrm{H}-1)\frac{r^2}{a^2}\right].\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

We see that $\sigma_l \gtrless \sigma_c$ according as $H \gtrless 1$. Continuing as before, we find successively that

$$\theta = \frac{2hr}{a^2}\left[1+\left\{(H-\tfrac{1}{3})\frac{r^2}{a^2}-\tfrac{1}{2}(H+1)\right\}\frac{h^2}{a^2}\right], \quad . \quad . \quad . \quad . \quad . \quad (22)$$

$$\frac{z}{h}=\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{2}(H+1)\frac{r^2h^2}{a^4}\right], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

$$\frac{p}{Y}=\frac{4ht_0}{a^2}\left[1+\tfrac{1}{2}(3H-5)\frac{h^2}{a^2}\right], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

$$v=\frac{2hr}{a^2}\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{2}\left\{(7H-5)\frac{r^2}{a^2}-(3H-1)\right\}\frac{h^2}{a^2}\right], \quad . \quad (25)$$

$$\frac{u}{a}=\frac{h^2r}{a^3}\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{4}\left\{(7H+1)\frac{r^2}{a^2}-(3H+1)\right\}\frac{h^2}{a^2}\right], \quad . \quad (26)$$

$$\ln\frac{t_0}{t}=\frac{2h^2}{a^2}\left(1-\frac{r^2}{a^2}\right)\left[1+\tfrac{1}{4}\{(13H-7)\frac{r^2}{a^2}-(3H-1)\}\frac{h^2}{a^2}\right]. \quad . \quad (27)$$

The polar radius of curvature is given by

$$\frac{a}{\rho}=\frac{2h}{a}\left[1-\tfrac{1}{2}(H+1)\frac{h^2}{a^2}\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

The equation of a circle passing through the edges and the pole is (to this approximation)

$$\frac{z'}{h}=\left(1-\frac{r^2}{a^2}\right)\left(1+\frac{r^2h^2}{a^4}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

On comparing (23) and (29), it is found that the maximum relative difference between the bulge profile and the circle is

$$\text{Max.}\left(\frac{z-z'}{h}\right)=\tfrac{1}{8}(H-1)\frac{h^2}{a^2}, \text{ when } r=\frac{a}{\sqrt{2}}.$$

This vanishes when $H=1$, in agreement with the solution for the special stress–strain curve (12), since $\exp\epsilon \sim 1+\epsilon$ to this approximation. When $H>1$ (which is so for all save exceptional degrees of cold-work) the profile of the bulge lies above the circle (29), and below when $H<1$. Expressed otherwise, the polar circle of curvature cuts the original plane outside the perimeter of the die when $H>1$ and inside when $H<1$. This was found experimentally by Brown and co-workers at all stages up to the pressure maximum ($H\sim\frac{1}{2}$ for the hard copper used by Brown and Sachs and ~ 4 for the hard aluminium alloy used by Brown and Thompson).

From (27) the polar strain is

$$\epsilon=\frac{2h^2}{a^2}\left[1-\tfrac{1}{4}(3H-1)\frac{h^2}{a^2}\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Combining this with (28) we obtain

$$\frac{\rho}{a}=\left[\frac{1}{2\epsilon}+\frac{(H+5)}{16}\right]^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

Thus, the effect of the rate of work-hardening on the $(\rho/a, \epsilon)$ relation is comparatively slight up to strains of order 0·3, because of the square root dependence and the small contribution made by the term H/16. (When H=1, equation (31) is the second approximation to the solution of (15), as is to be expected.) The relation between bulge height and polar strain, or between bulge height and polar curvature, is much more dependent on the value of H. We see from (30) that, for a given height, the most severely worked metals (small H) have the largest polar strains but the smallest polar curvatures; this is in agreement with the observations of Brown and Thompson.

An inspection of equation (28) indicates that a rough criterion for the range of validity of this solution (say to an accuracy of 1 per cent) is

$$(\mathrm{H}+1)\frac{h^2}{a^2}<\tfrac{1}{2}.$$

Thus, when H=1, the solution is probably reliable up to a polar height of $\frac{1}{2}a$. When H=3, a value typical of moderately cold-worked metals, the solution is probably reliable up to a polar height of $a/2\sqrt{2}\sim 0{\cdot}35a$. It is only when H is less than about $\frac{1}{2}$ that the range of validity probably extends past the pressure maximum which, according to (24), occurs at a polar height of

$$\left[\frac{2}{3(5-3\mathrm{H})}\right]^{1/2}a.$$

Thus, the solution does not appear capable of explaining the secondary bulge observed by Brown and co-workers; this may indeed be peculiar to annealed metals with a well-rounded stress–strain curve. It has not yet been found possible to obtain any simple approximate solution for such metals.

ACKNOWLEDGMENTS.

I wish to thank the Director of the British Iron and Steel Research Association for permission to publish this work, and Mr. I. M. Longman, of the B.I.S.R.A. Solid Mechanics Section for checking the accuracy of the series solution.

REFERENCES.

BROWN, W. F., JR., and THOMPSON, F. C., 1949, *Trans. Am. Soc. Mech. Eng.*, **71,** 575.

BROWN, W. F., Jr., and SACHS, G., 1948, *Trans. Am. Soc. Mech. Eng.*, **70,** 241.

GLEYZAL, A. N., 1948, *Trans. Am. Soc. Mech. Engrs.*, **70,** 288.

LANKFORD, W. T., and SAIBEL, E., 1947, *Trans. Am. Inst. Min. Met. Eng., Metals Division*, **171,** 562.

CI. *A Method for the Mapping of Vector Potential Distributions in Axially Symmetrical Systems.*

By G. LIEBMANN,
Associated Electrical Industries Research Laboratory, Aldermaston, Berks.*

[Received July 27, 1950].

SUMMARY.

The magnetic vector potential distribution in axially symmetrical systems containing current carrying conductors and iron of arbitrary shape can be obtained experimentally with good accuracy using resistor networks. Two networks are possible, one for plotting the flux function rA_θ and one for plotting the vector potential A_θ itself. The latter one is a modification of a field plotting network described previously. Some confirming measurements on a model of an infinitely long coil were carried out.

§1. INTRODUCTION.

THE determination of the magnetic field due to currents flowing in conductors requires the mapping of the magnetic vector potential. In the presence of iron a mathematical analysis becomes almost impossible, but it was shown by Pérès and Malavard (1938) and by Peierls (1946) that the problem can be solved experimentally with the help of the electrolytic tank analogue in those cases where all currents flow in straight parallel conductors. The problem is then a two-dimensional one, and only one component (*e.g.* the z-component) of the vector potential exists, which satisfies a Poisson equation in x and y. In the electrolytic tank analogue the iron circuits are represented by wax models; the current conductors, supposed to be perpendicular to the surface of the electrolyte, are represented by a number of probes dipping into the electrolyte, and feeding the appropriate currents into the tank.

This model representation cannot be applied to rotationally symmetrical problems, where all currents flow in circles concentric with the axis of the system. The vector potential has then only an angular component A_θ which has to satisfy this equation:

$$\frac{\partial^2 A_\theta}{\partial z^2}+\frac{\partial^2 A_\theta}{\partial r^2}+\frac{1}{r}\frac{\partial A_\theta}{\partial r}-\frac{1}{r^2}A_\theta=-\frac{4\pi}{10}J, \quad . \quad . \quad . \quad . \quad (1)$$

where J is the current density in the conductors, in amps/cm.^{2}.

Writing $$rA_\theta=\Psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

* Communicated by the Author.

this equation can be put into the form

$$\frac{\partial^2\Psi}{\partial z^2}+\frac{\partial^2\Psi}{\partial r^2}-\frac{1}{r}\frac{\partial\Psi}{\partial r}=-\frac{4\pi}{10}r\mathrm{J}, \quad \ldots\ldots \quad (3)$$

It was recently shown by Peierls and Skyrme (1949) that an experimental solution for this function Ψ can be obtained by using two tank measurements, first setting up the model in a tank with equal height of electrolyte, and then in a tilted or wedge tank. The difference between the two measurements can be used as a correction for the results obtained in the first tank, provided the correction is small, *i.e.* measurements are taken not too near the axis.

§2. APPLICATION OF RESISTOR NETWORK METHOD.

A *direct* measurement either of the vector potential A_θ, satisfying equation (1), or of the function (rA_θ), satisfying equation (3), can be obtained with a resistor network similar to the type described recently by the present author.

Consider the equation

$$\frac{\partial^2\mathrm{F}}{\partial x^2}+\frac{\partial^2\mathrm{F}}{\partial y^2}=\mathrm{G}(\mathrm{F}, x, y). \quad \ldots\ldots\ldots \quad (4)$$

In the earlier paper (Liebmann 1950) it had been shown that the difference equation corresponding to equation (4) is satisfied by a network of interconnected resistances of equal value R_0. The values of the function F are represented by the voltages at the resistor junctions, and the function G is related to the currents I, fed into the junctions from an external source, through the equation

$$\mathrm{G}=-\mathrm{I}R_0/(\Delta x)^2. \quad \ldots\ldots\ldots\ldots \quad (5)$$

Now let us apply the transformation

$$\left.\begin{aligned} x&=r^{-k}z \\ y&=\frac{r^{1-k}}{1-k}\end{aligned}\right\} k\neq 1, \qquad \left.\begin{aligned} x&=z/r \\ y&=\log r\end{aligned}\right\} k=1, \quad \ldots\ldots \quad (6)$$

where k is any real constant. This transforms the functions $\mathrm{F}(x, y)$ and $\mathrm{G}(\mathrm{F}, x, y)$ into $f(r, z)$ and $g(f, r, z)$ and equation (4) into

$$\frac{\partial^2 f}{\partial z^2}+\frac{\partial^2 f}{\partial r^2}+\frac{k}{r}\frac{\partial f}{\partial r}=r^{-2k}\,g(f, r, z), \quad \ldots\ldots\ldots \quad (7)$$

or with

$$r^{-2k}g=g^* \quad \ldots\ldots\ldots\ldots\ldots \quad (8)$$

into

$$\frac{\partial^2 f}{\partial z^2}+\frac{\partial^2 f}{\partial r^2}+\frac{k}{r}\frac{\partial f}{\partial r}=g^*. \quad \ldots\ldots\ldots \quad (9)$$

If we now choose the four resistances $R_1 \ldots R_4$, joined in the network junction P_0 (see fig. 1) such that

$$\left.\begin{aligned} R_1&=R_3=r^{-k}R_0, \\ R_2&=(r_0^{1-k}-r_2^{1-k})R_0/(1-k), \\ R_4&=r_4^{1-k}-r_0^{1-k})R_0/(1-k), \end{aligned}\right\} \quad \ldots \ldots \quad (10)$$

Fig. 1.

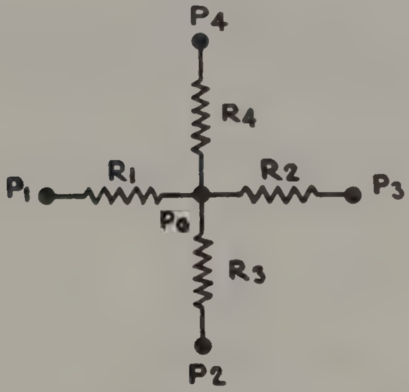
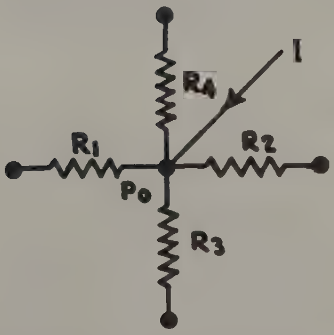


Fig. 2.



R_0 being an apparatus constant, we obtain a network which solves the differential equation (7) to a higher degree of approximation than the difference equation which would replace equation (7) in numerical methods of solution †.

---

† Except for $k \rightarrow 0$, the difference equation as used in most numerical techniques is only correct to second order terms, whereas the specified network is correct to third order terms, whether $k \rightarrow 0$ or not. This point is mostly not very important in regions distant from the axis, but cannot be overlooked as the axis is approached.

In the previous paper (Liebmann 1950) it was shown, through the application of Kirchhoff's law, that if the function F is represented by voltages, then the function G is represented by currents fed into the network points P_0 from an external source (see fig. 2). Similarly, we find that the function g^* is related to the currents I fed in at P_0 by

$$I=-g^*(\Delta z)^2/R_1, \quad \ldots \ldots \ldots \ldots \quad (11$$

where R_1 is given by equation (10).

§3. RESISTOR NETWORK FOR FINDING (rA_θ).

The two alternative equations (1) and (3) for the vector potential A_θ and for (rA_θ) can therefore be solved by constructing the appropriate networks.

Taking first equation (3), we have $k=-1$, and

$$\left.\begin{aligned} R_1&=R_3=rR_0,\\ R_2&=(r_0{}^2-r_2{}^2)R_0/2,\\ R_3&=(r_4{}^2-r_0{}^2)R_0/2. \end{aligned}\right\} \quad \ldots \ldots \quad (12)$$

The currents to be fed in are then

$$I_0=+\frac{4\pi}{10}J(\Delta z)^2/R_0. \quad \ldots \ldots \ldots \ldots \quad (13)$$

Obviously, current would only be fed in at the positions occupied by the conductors. If iron is present, the network must be open-circuited along the boundaries of the iron, the network being correctly terminated there, as discussed in the earlier paper, in order to fulfil the boundary condition $\frac{\partial(rA_\theta)}{\partial n}=0.$

§4. RESISTOR NETWORK FOR FINDING A_θ.

It will mostly be more attractive to use a network which solves directly for A_θ. Then $k=+1$, and the previously described network, representing the Laplacian operator for axially symmetrical fields, can be applied. The currents to be fed-in would then have to be $\left(\frac{4\pi}{10}J-\frac{1}{r^2}A_\theta\right)\Delta z^2/R_1$; this would require the use of the iteration method described before (Liebmann 1949). This need is eliminated by the following modification: Each network junction is connected through an external resistance of value $R_5=rR_0$ to a point of zero potential (fig. 3).

We then have these values of resistances :

$$\left.\begin{aligned} R_1&=R_3=R_0/r_0,\\ R_2&=R_0\log(r_0/r_2),\\ R_4&=R_0\log(r_4/r_0,\\ R_5&=r_0R_0. \end{aligned}\right\} \quad \ldots \ldots \quad (14)$$

This network now represents the operator

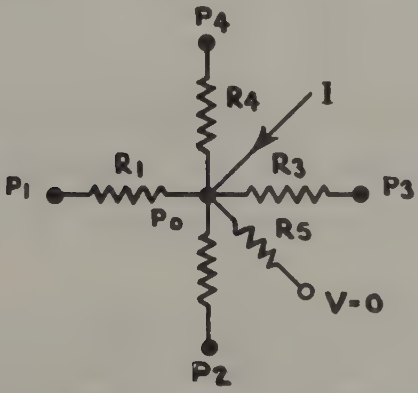
$$\left[\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}\right]$$

which differs from zero only at the position of current carrying conductors. at these positions, the current

$$\mathrm{I}=\frac{4\pi}{10}\mathrm{J}\ (\Delta z)^2/\mathrm{R}_1=\frac{4\pi}{10}\mathrm{J}r(\Delta z)^2/\mathrm{R}_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

is fed into the network. It is thus possible to measure the vector potential distribution inside current carrying regions within the limitations imposed by the finite size of the interval. Iron circuits are again represented by open circuiting the appropriate boundaries, as $\partial \mathrm{A}_\theta/\partial n=0$, for infinite permeability; the values of the resistances R_5 have to be doubled along the boundaries and quadrupled at interior corners to maintain correct relations at the boundaries. If it is desired to represent correctly a finite value of permeability, the network constants have to be modified, by increasing the value of the constant R_0 in the ratio of $(\mu_{\mathrm{iron}}/\mu_{\mathrm{air}})$, in all resistances within the area occupied by the iron, or by tapping down the voltage in the ratio of $1:\mu$ at the boundary points, using the resistances R_5 as potential dividers, in order to satisfy the boundary condition $\mathrm{A}_{\theta\,\mathrm{iron}}=1/\mu\ \mathrm{A}_{\theta\,\mathrm{air}}$. As previously described, curved boundaries can be simulated by shunting the network resistances; the values of the resistors R_5 have then to be adjusted in a similar way, as mentioned before when referring to the boundaries of iron circuits, taking the change of size of the effective area of the network element $(\Delta z\times\Delta r)$ into account.

Fig. 3.



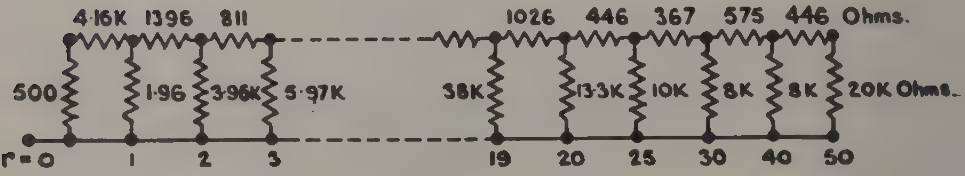
If a problem to be solved involves regions which are open in the direction away from the axis, then it will be necessary to follow the technique applied in similar cases of electrostatic fields. A small scale model is set up on the network, in which the distance of the current carrying conductors from the axis is small compared with the radial extension of the network, and the approximate values of the vector potential at a suitable radial distance r_B are determined. The model is then enlarged so that this relative radial value r_B now corresponds to the edge of the network, and the potential values required at this distance from the axis are forced on to the model by feeding in additional currents at these points. Alternatively, a suitable flux line may be made an equipotential for the voltage. The axis is, of course, an equipotential line of zero potential.

Instead of determining the values of A_θ at the mesh points, one can plot out directly the lines of equal A_θ using the ganged potentiometer arrangement described before (Liebmann 1950).

§5. PRACTICAL REALIZATION AND TESTS OF ACCURACY.

The correctness of the discussed method was confirmed by several experiments with a linear resistance strip. This strip (fig. 4) represented one section of a precision (r, z) resistance network being built in this laboratory. The gradation of the resistance values was on a logarithmic basis (equation (14)), and the values were the same as those employed in the previously described resistance network ($R_0=2000$ ohms, except that the resistance values were kept within about $\pm 0{\cdot}2$ per cent of their nominal values, and that the termination strip was extended to go out from $r=20$, the edge of the network proper, to $r=50$ by a few large steps. Each junction, at distance r from the axis, was connected to the axis $(V=0)$ by a resistor of value $R_5=rR_0$, according to equation (14). The finite size of the interval makes the nominal value rR_0 incorrect by a small factor, of the order of 1·8 per cent at $r=1$ and 1·2 per cent at $r=2$, the correction factor diminishing to a negligible amount for greater values of r. Expressing the length r through its equivalent resistance value leads to the corrected value $R'_5=\frac{1}{4}R_0^2\sum_1^4(1/R_i)$; these corrected values were used in the final experiments. The linear resistance strip can be used to investigate problems in which A_θ does not depend on z, *e.g.* the case of infinitely long coils of constant radii. In one such experiment a single layer coil of $r_0=10$ units was assumed, the potential being set to the relative value $A_0=1$ at $r_0=10$. Table I. gives the experimentally found values and the theoretical

Fig. 4.

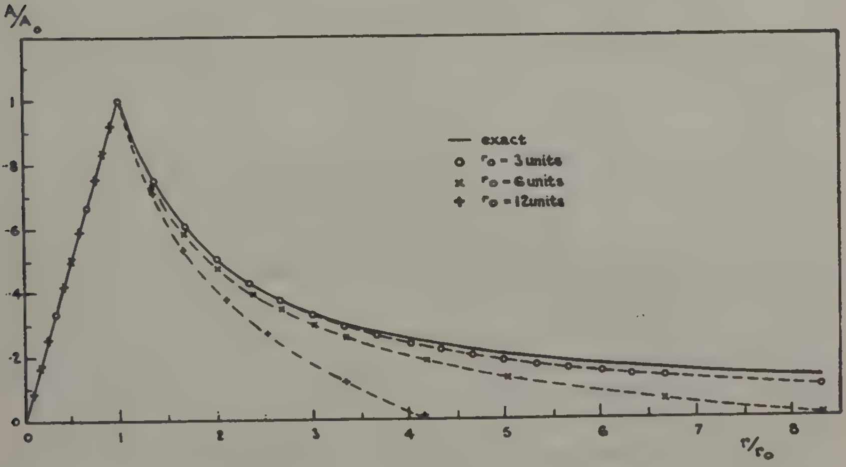


values $A_\theta = A_0 r/r_0$. The agreement between experimental values and theoretical values is very good, particularly in view of the small number of components involved, which does not bring the averaging properties of the network fully into play.

TABLE I.

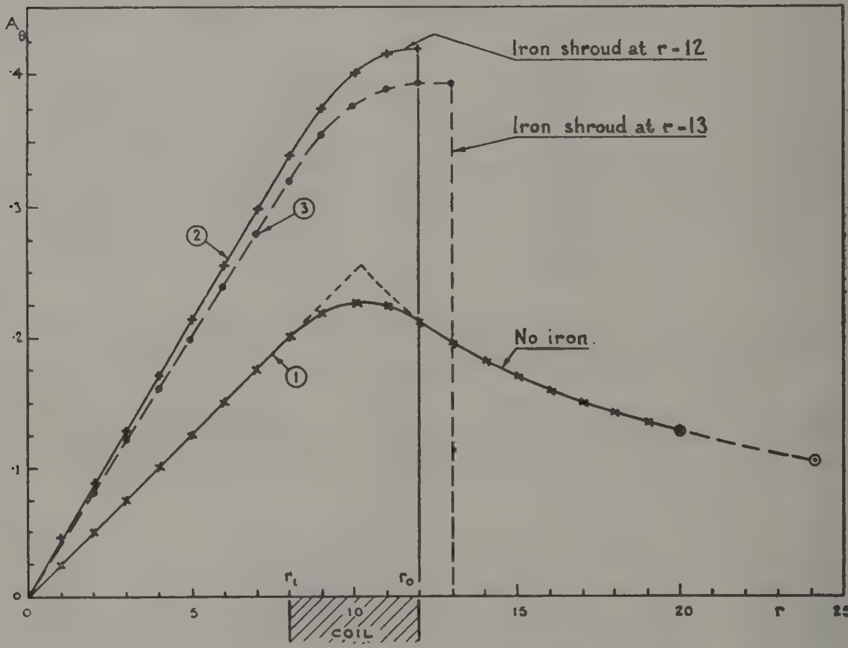
r/r_0	$A_\theta(r)$ exp.		A(r) calc.
	Before correction, $R_5 = rR_0$.	After correction, $R'_5 = \frac{1}{4}R_0^2\sum_1^4(1/R_i)$	
0	0	0	0
0·1	0·1002	0·1000	0·1000
0·2	0·2012	0·2004	0·2000
0·3	0·3012	0·3004	0·3000
0·4	0·4011	0·4004	0·4000
0·5	0·5009	0·5002	0·5000
0·6	0·6007	0·6002	0·6000
0·7	0·7005	0·7001	0·7000
0·8	0·8003	0·8001	0·8000
0·9	0·9002	0·9001	0·9000
1	1	1	1

Fig. 5.



The vector potential A_θ falls away as $1/r$ outside an infinitely long single layer coil, and an incorrect field distribution will be obtained at some distance from the coil unless the model is made very small in relation to the network, giving low " resolution " near the coil itself. This difficulty can be eliminated by first measuring the field distribution of a small scale model in which the potential $A_\theta = 0$ at $r = 50$ units, and by then setting up the full size model and impressing the potential values found in the first test on the new boundary points. To study the quantitative side of this procedure the field distributions outside single layer coils of radii $r_0=3$, $r_0=6$ and $r_0=12$ mesh units were investigated. The results are plotted in fig. 5, where the full curve represents the theoretical curve $A_\theta(r)=A_0 r_0/r$. For the smallest coil, good agreement between theoretical and experimental values (with $A_\theta=0$ at $r=50$) is found up to about $r/r_0 \simeq 2{\cdot}5$. The error is here $-0{\cdot}004\ A_0$. From graphs of the kind given in fig. 5 information about the appropriate size of the small scale model for the first test can be obtained.

Fig. 6.



The effect of an iron shroud on a coil of inner radius $r_i=8$ cm. and outer radius $r_0=12$ cm. is shown in fig. 6. Curve 1 gives a plot of $A_\theta(r)$ for a coil without iron being present. The boundary at $r=20$ was set to the value obtained from a quarter scale model as described before (circled points). The total current fed into the network was $0{\cdot}2500\ V_0$ ma, where V_0 is the feeding voltage of the network *), or $J=6{\cdot}25\times10^{-5}\times V_0$ amps/cm.².

Within the coil, the measurements give $A_\theta=0{\cdot}02508\ rV_0$ gilberts, hence $H_z=1/r\ \partial(rA_\theta)/\partial r=0{\cdot}0501_6\ V_0$ oersteds. Calculating H_z direct from the formula $H_z=4\pi/10$ AT/cm. we obtain $0{\cdot}0500\ V_0$ oersteds; this appears to be a quite satisfactory agreement with the experimental results †.

---

* All measurements of voltages and currents were referred to resistance measurements, as this gives a higher accuracy and greater convenience; the feeding voltage V_0 appears then as a common multiplication factor.

† The continuations of the inner and outer branches of the A_θ curve, which obey the (kr) and (k/r) relationships, would intersect at $r=10{\cdot}2$, the centre of gravity of the coil windings.

Placing an iron shroud ($\mu = \infty$) tightly around the coil (at $r=12$) changes the shape of the A_θ curve within the coil considerably (curve 2, fig. 6), and increases the value of A_θ inside the coil to $A_\theta = 0{\cdot}04243\, rV_0$, *i.e.* the iron shroud has increased the slope of the A_θ curve, and therefore the magnetic field within the coil, by a factor $1{\cdot}69_2$. If the iron shroud is placed a small distance away from the coil, at $r=13$ (fig. 6, curve 3) the A_θ curve becomes modified, and the factor of increase in field strength in only $1{\cdot}60_0$. A finite permeability ($\mu \simeq 1000$) was then simulated, which produced very slight modifications of the A_θ curve near its maximum; the change is too small to be represented in the graph.

§ 6. CONCLUSION.

It is seen from these tests that a satisfactory accuracy can be obtained. As the modifications to existing field plotting networks, satisfying Laplace's equation, can be easily effected, the measurement of vector potential distributions becomes a very practicable proposition, even within current carrying regions and in the presence of iron. The method should therefore have applications not only in the design of betatron and synchrotron magnets, and in similar problems, but also in many ordinary problems of electrical machine design. The same principle can also be applied to the mapping of other vortex fields of rotational symmetry, when the vorticities and positions of the vortex rings are given.

ACKNOWLEDGMENT.

The author is indebted to Dr. T. E. Allibone, F.R.S., for permission to publish this paper.

REFERENCES.

LIEBMANN, G., 1949, *Nature, Lond.*, **164,** 149 ; 1950 *Brit. Journ. Appl. Phys.*, **1,** 92.
PEIERLS, R. E., 1946, *Nature, Lond.* **158,** 831.
PEIERLS, R. E., and SKYRME, T. H. R., 1949, *Phil. Mag.*, **40,** 249.
PÉRÈS, J., and MALAVARD, L., 1938, *Bull. Soc. Franç. Elect.*, **8,** 715.

CII. *Double Stars due to Cosmic Rays.*

By T. T. Li,
Imperial College, London*.

Summary.

Cosmic ray stars are distributed at random over a photographic plate when distances exceeding 1 cm. are considered. For small distances of the order of 1 mm. the number of close pairs exceeds that expected on a random distribution. The effect is not found on plates exposed at sea-level and there is some evidence that it is more marked if dense matter surrounds the plates. Several mechanisms are suggested to account for the effect but none is free from objection.

§1. Introduction.

Measurements were made (Li and Perkins 1948) on the surface distribution of 1230 stars on 54 in.^{2} of 100μ C2 plates exposed for forty days at 3650 m. altitude. The plates were divided into $(1\times\frac{1}{2})$ in. areas to facilitate searching. It was found that the distribution of star frequencies between these areas conformed to the Poisson distribution, showing that the stars were randomly distributed over areas of these dimensions.

At smaller distances, however, the stars showed a tendency to be in close proximity of each other more frequently than would be expected if they were distributed at random. (See also Leprince-Ringuet and Heidmann 1948.)

Thus, for a random distribution, the number of close pairs of stars, separated by a projected distance in the plane of the emulsion $\leqq r_p$, should be

$$Q(r_p)=\sum_i \frac{Ni(Ni-1)}{2}\cdot\frac{\pi r_p^{\,2}}{ab},$$

where

Ni = number of stars in the ith area,

a, b = sides of the areas = 1 in. and $(\frac{1}{2})$ in. respectively, and r_p is small compared with b.

The actual observed number of close pairs, $P(r_p)$, was compared with the corresponding $Q(r_p)$. It was found that $P(r_p)$ was always greater than $P(r_p)$ up to $r_p=0{\cdot}03$ in., which was then the largest distance measured. Moreover, the ratio $(P-Q)/\sqrt{Q}$ of the observed to the expected deviations was always greater than one. Different mechanisms were suggested to account for this excess of close pairs, which was termed " double stars ".

---

* Communicated by Sir George Thomson, F.R.S.

In the present paper, the results are given of further measurements, extended to larger distances r_p and made on plates exposed under various conditions.

§2. EXPERIMENTAL RESULTS.

1. Measurements were made on the following sets of plates :

(*a*) Plates exposed at 3650 m. altitude without any absorbing material immediately above the plates. (A total of 1230 stars on 54 in.^{2}.)

(*b*) Plates exposed at sea-level with no absorbing material immediately above the plates. (140 stars on 27 in.^{2}.)

(*c*) Five boxes of plates exposed in a narrow crevasse ($\sim$10 cm. wide) in a glacier at 3650 m. altitude, at the following depths below the surface of ice : 0 m., 1 m., 2 m., 3·5 m., and 5·5 m. (1126 stars on 77·5 in.^{2}).

(*d*) Four boxes of plates exposed at 3650 m. altitude with the following thicknesses of lead above them : 0 cm., 5 cm., 15 cm., and 35 cm. (374 stars on 16 in.^{2}).

The plates were all 100μ thick of the Ilford C2 type. For the sake of convenience, they will be referred to as the " 3650 ", " sea-level ", " ice ", and " Pb " plates, respectively.

2. Measurements were extended to include all close pairs with a projected separation of $r_p \leqslant (0{\cdot}1)$ in. The formula in (1) for the expected number of close pairs from a random distribution is no longer valid, since r_p is comparable with b now. The following exact formula can be derived:

$$Q(r_p)=\sum_i \frac{Ni(Ni-1)}{2}\cdot\frac{\pi r_p^2}{ab}\cdot\left[1-\frac{4(a+b)r_p}{3\pi ab}+\frac{r_p^2}{2\pi ab}\right]$$

which holds true for $0 \leqq r_p \leqq b$.

In the following table, the values of Q, P, $(\Delta P-\Delta Q)^2/\Delta Q$, and $(P-Q)/(\Sigma Ni)$ are tabulated, using the data of all four sets of plates :

r_p, (1/100) in.	2	3	4	5	6	7	8	9	10
$P(r_p)$	46	107·5	182·5	280	239	505	640	791	936
$Q(r_p)$	37·6	82·2	159·1	230·6	329·2	444·1	573·1	717·7	876·5
$(\Delta P-\Delta Q)^2/\Delta Q$	1·9	6·4	0·0	9·5	1·8	0·0	0·3	0·3	1·2
$10^3\,(P-Q)/(\Sigma Ni)$	2·9	8·8	8·2	17·2	21·2	21·2	23·3	25·6	20·8

The last row shows the excess of close pairs per 1000 stars. There seems to be a saturation after $r_p \sim (6/100)$ in. Whether this is real or whether it simply arises from the fact that $Q(r_p)$ increases as $\sim r_p^2$ while $(P-Q)$ increases more slowly with r_p so that the effect gets relatively small at large distances and is lost in the statistical fluctuations, it is not possible to say. (Also $(P-Q)$ must be equal to zero at $r_p=a$, since, by definition, $Q_{r_p=a}=\Sigma Ni(Ni-1)/2=P_{r_p=a}$).

By applying the χ^2-test, it is found that the probability that this excess is only the result of statistical fluctuations is small, $\sim$0·5 per cent. Hence we can conclude that this excess does exist.

3. By comparing results of the different sets of plates, we found that

(*a*) There is no excess of P over Q in the " sea-level " plates, or else the effect is so small that it is outweighed by the statistical fluctuations.

r_p, (1/100) in.	2	3	4	5	6	7	8	9	10
10^3 (P Q)/(ΣN*i*)	−2·9	7·2	2·1	−5·0	−12·9	6·4	2·1	−8·6	−7·2

It is seen from the above figures that the values of (P−Q)/(ΣN*i*) fluctuate between positive and negative. And the χ^2-test gives a probability of 10 per cent for these values of (P−Q) to be due to statistical fluctuations.

(*b*) If we group together (i) all the plates which were exposed at 3650 m. with no absorbing material immediately above them (*i. e.* the " 3650 " plates, the first box of the " ice " plates, and the first box of the " Pb " plates), and (ii) those which were exposed under a thickness of ice or lead (*i. e.* boxes 2 to 5 of the " ice " plates and boxes 2 to 4 of the " Pb " plates), we notice that the values of (P−Q)/(ΣN*i*) for the latter group are greater than the corresponding ones of the former group :

r_p, (1/100) in.		2	3	4	5	6	7	8	9	10
10^3: (P−Q)/	−(ii)	4·4	12·5	14·0	20·0	30·0	24·4	38·6	37·7	38·7
(ΣN*i*)	−(i)	2·5	6·8	11·0	17·3	19·9	14·9	16·0	21·2	11·8

But the probability that the difference is due to statistical fluctuations is 25 per cent, which is not small enough to enable us to establish the existence of a difference.

§3. Discussions.

1. Possible Errors : Let us examine the possible errors from which an apparent excess (P−Q) may arise.

(*a*) *Edge effect.*—The damage of the emulsion near the edge of the plates and at the ruled lines of the areas reduces the effective area of each rectangle to less than half sq. in. by an unknown amount. This might affect the calculation of Q to a significant extent. To eliminate this, let us study only the stars inside an imaginary rectangle of 1 cm. by 2 cm. at the centre of each area. The coordinates of the stars enable us to exclude all stars outside this central rectangle, thus giving a true area of 2 cm.^{2} without any error due to incorrect ruling and damaged emulsion at the edge.

We took, as a random sample, the plates exposed under 0 m., 1 m., and 2 m. of ice. There are 52 areas altogether. The distribution of stars in these 52 1-cm. by 2-cm. rectangles is as follows :

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	17
number of 1×2 cm.^{2} areas with *n* stars	1	1	0	1	4	6	6	8	8	5	5	2	2	2	1

Thus, considering the central 1 cm. by 2 cm. areas only, we have

$$\Sigma \mathrm{N}i(\mathrm{N}i-1)/2=(1\times3\times2+4\times4\times3+\ .\ .\ .\ +2\times13\times12+1\times17\times16)/2$$

$$=1508.$$

$$\therefore\quad \mathrm{Q}(r_p)=1508\,.\frac{\pi r_p^{\,2}}{ab}\left[1-\frac{4(a+b)r_p}{3\pi ab}+\frac{2\pi ab}{r_p^{\,2}}\right].$$

Here $a=2$ cm. and $b=1$ cm. By comparing these expected numbers of close pairs with the observed number inside the central areas, we get the values of (P−Q)/(ΣNi):

r_p, (1/100) in.	2	3	4	5	6	7	8	9	10
$10^3\,.\,(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$	8·0	23·0	21·0	15·4	29·8	23·0	11·3	12·8	−14·1

The values of (P−Q)/(ΣNi) for the same 52 areas but considering the whole (1×½) in. area are given below:

4·3 12·8 15·7 16·1 30·2 24·8 28·2 26·8 13·0

A comparison of the two sets of values shows that the "Edge effect" does not contribute much, if any, to the excess (P−Q).

(*b*) *Uneven Thickness of Emulsion.* Uneven thickness of the emulsion has an effect of concentrating stars and thus giving rise to a spurious excess, (P−Q). But it can be shown that even for a variation of thickness of up to 20 per cent, the effect is still very small.

First consider a simple case. Suppose in the area (1×½) in., half of the emulsion has a thickness d, and the other half has a thickness d', $d/d'=1{\cdot}2$. Let n and n' be the numbers of stars in the two halves respectively. Then $n/n'=d/d'=1{\cdot}2$.

If the stars were randomly distributed, then the number of close pairs that we should observe in the two halves is

$$\mathrm{P}(r_p)=\frac{n^2}{2}\cdot\frac{\pi r_p^{\,2}}{s}+\frac{n'^2}{2}\cdot\frac{\pi r_p^{\,2}}{s},$$

for small values of r_p, s being $(\frac{1}{2})\times(\frac{1}{2})$ in.^{2}.

On the assumption that the emulsion thickness is uniform, we calculate the expected number of close pairs in the whole area as

$$\mathrm{Q}(r_p)=\frac{(n+n')^2}{2}\cdot\frac{\pi r_p^{\,2}}{2s}.$$

Hence the spurious excess of P over Q due to non-uniform thickness of the emulsion is

$$\mathrm{P}-\mathrm{Q}=\frac{\pi r_p^{\,2}}{4s}[2n^2+2n'^2-(n+n')^2]=\frac{\pi r_p^{\,2}}{4s}\,.\,(n-n')^2$$

and

$$(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)=(\mathrm{P}-\mathrm{Q})/(n+n')=\pi r_p^{\,2}\,.\,(n-n')^2/4s(n+n').$$

Inserting $n/n'=1\cdot 2$, we get

$$(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)=\frac{\pi r_p^2}{4s}\cdot\frac{0\cdot 04n'}{2\cdot 2}.$$

Assuming a value of $n'=6$, say, we get

r_p, (1/100) in.	2	3	4	etc.
$10^3\,.\,(\mathrm{P}-\mathrm{Q})/\Sigma\mathrm{N}i)$	0·14	0·31	0·56	etc.

These numbers are negligible compared with the observed values of $(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$. If the emulsion consists of two unequal areas of different thicknesses, the spurious excess can be shown in a similar way to be smaller still. Hence we conclude that the contribution due to non-uniform thickness of the emulsion is negligible.

(*c*) *Stars missed owing to Inefficient Searching.*—Inefficiency in searching might also give rise to a spurious excess of P over Q. For instance, if most of the stars missed are those which are at a large distance to any other star, then a spurious (P—Q) can result.

The efficiency of searching depends on the time spent on an area. At the rate (1 in.^{2}/day) with which the plates were searched, we found that about 20 per cent of the stars were missed. The " ice " plates have been searched twice, first at the normal speed and then at half that speed. The overall efficiency thus obtained was 99 per cent.

Assume the extreme case in which all the stars missed were those that were at a distance >(1/10) in. from any other star, so that they would increase the values of $(\mathrm{Q}(r_p)$ but not $\mathrm{P}(r_p)$. Assume a conservative efficiency of searching, in the case of the " ice " plates, of only 98 per cent.

$\bar{n}$=Average number of stars/area=1126/155=7·26.
$\bar{n}'$=Average number of stars/area, corrected for inefficient searching=7·26/0·98=7·41.

$\bar{n}'(\bar{n}'-1)/\bar{n}(\bar{n}-1)=1\cdot 025.$

$\mathrm{Q}'(r_p)$=corrected values of $\mathrm{Q}(r_p)=1\cdot 025\ (\mathrm{Q}r_p)$.

We have

r_p. (1/100) in.	2	3	4	5	6	7	8	9	10
$\mathrm{P}(r_p)$	17	43·5	82	127	182·5	229	294	352	416
$\mathrm{Q}'(r_p)$	17·3	38·5	67·5	103·9	147·8	198·2	255·3	318·7	388·1
$10^3\,.\,(\mathrm{P}-\mathrm{Q})/(\Sigma\mathrm{N}i)$	—·3	4·3	12·6	20·0	30·1	26·7	33·6	28·9	24·1

We see that after correction, assuming the extreme case, we are still left with large values of $(\mathrm{P}-\mathrm{Q})/(\Sigma \mathrm{N}i)$. Moreover, if the inefficiency of searching were the cause of the excess, (P—Q), then there should be a vast difference between the results of the "ice" plates and the other plates—the values of $(\mathrm{P}-\mathrm{Q})\ (\Sigma \mathrm{N}i)$ for the " ice " plates would be much smaller than those for the rest of the plates. This difference was not found. This shows that the inefficiency in searching does not contribute much to the values of (P—Q).

2. In this section we give some calculations of the possible causes of the " double stars " and compare them with the experimental data.

(*a*) *Regenerative process*—If we assume that the double stars are produced by penetrating radiation, emitted isotropically from the primary stars, producing secondary stars, then the number of double stars with separation $\leqslant r$ is (Li and Perkins 1948).

$$\mathrm{R}=(\Sigma \mathrm{N}i)\, q\, \sigma\, \mathrm{L}\, [3d/4+d/2 \log_0 (r/d)],$$

for $r>d$, neglecting absorption.

Here d = thickness of the emulsion $=100\mu$,
L = number of nuclei (excluding hydrogen) per c.c. of emulsion $=5{\cdot}6\times 10^{22}$,
q = number of penetrating particles from each star,
σ = cross-section of production of secondary stars by the penetrating particles.

Since the values of r we are considering are $\geqslant d$, we can take $r=r_p$. By comparing the calculated values of $\mathrm{R}/(\Sigma \mathrm{N}i)$ with the experimental values of $(\mathrm{P}-\mathrm{Q})/\Sigma \mathrm{N}i)$, we get

$$q\sigma=(13{\cdot}8\pm 5{\cdot}4)\times 10^{-24}\ \mathrm{cm}^2.$$

A cross-section of the order of 10^{-24} cm. requires that q, the number of penetrating particles emitted per star, to be ~ 14. It seems almost impossible that there are so many particles emitted per star with energy large enough to produce another star.

The values of q for known particles can be estimated as follows: In recent experiments with Ilford G5 emulsion at 3560 m. altitude (Brown *et al.* 1949), 1223 fast particles (with specific ionization $\leqslant 1{\cdot}5$ times the minimum ionization of a singly charged particle) were found to be emitted from 4923 stars. Allowing for an equal number of fast neutral particles, we get a value of 0·5 fast particle per star. Taking the number of grey tracks (ionization between 1·5 to 5 times the maximum ionization of a singly charged particle) per star found for the star group with 7 and 8 heavily ionizing tracks as the average number of grey tracks per star for all stars, and multiplying by 2·25 to allow for neutrons, we get a value of 3 per star. Assuming that half of these have energies large enough to produce another star, we get a total of 2 particles emitted per star, capable of producing a secondary star. Examples of secondary stars produced by fast particles emitted from primary stars have actually been observed with electron sensitive plates. Not all the fast particles end in this way. The contribution they offer, if assumed to be emitted isotropically, can thus account for less than 2/13·8 or 15 per cent of the effect observed.

(*b*) Single penetrating particles producing one or more stars in their passage through the emulsion. In this process, all stars are assumed to be produced by penetrating particles capable of producing more than one star before they lose all or a large part of their energies. A certain directional distribution of the penetrating particles has to be assumed. An examination of a few particular cases will give us some idea of the cross-section required by such a process.

(i) If we assume that the penetrating particles are all coming in a direction perpendicular to the plane of the emulsion, it can easily be shown that the number of double stars is

$$R=(\Sigma Ni)\ d/2\lambda,$$

where λ=mean free path between successive productions of stars in the emulsion.

Taking 20/1000 as the value of $(P-Q)/(\Sigma Ni)$, we get $\lambda=2{\cdot}5$ mm. Or $\sigma=07\times10^{-24}$ cm$\frac{1}{3}$.

(ii) If we assume that the penetrating particles are all coming in a direction parallel to the plane of the emulsion, the number of double stars would be

$$R=(\Sigma Ni)\ r/\lambda,$$

giving a value of $\lambda=9{\cdot}3$ cm. or $\sigma=2\times10^{-24}$ cm^2.

Fig. 1*a*.

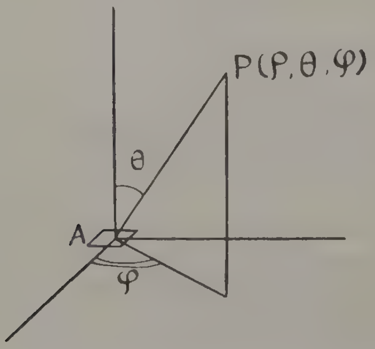
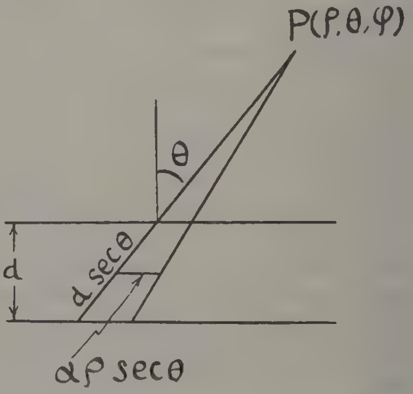


Fig. 1*b*.



(iii) If we assume that the penetrating particles are coming equally probably in all downward directions, then the expression for the number of double stars is simply the same as that for regenerative process except that q is replaced by unity, giving $\sigma=14\times10^{-24}$ cm^2..

Case (iii) is presumably closest to the actual directional distribution, so that with an assumed value of $\sigma=10^{-24}$ cm.^{2} this process could account about 7 per cent of the observed double stars.

(*c*) Stars produced by sheafs of particles. If we assume that the stars are produced by sheafs of particles, which are generated in the material surrounding the plates and are emitted isotropically in all downward directions, we can calculate the number of double stars produced.

Let K=number of sheafs generated per unit solid angle per unit volume per unit time,
α =their average conical angle,
S =average number of particles per sheaf,
T =total time of exposure,
λ =m. f. p. for star production in the emulsion,
λ' =m. f. p. for absorption in the surrounding material,
A =area of plates.

Then the number of stars produced in A is

$$(\Sigma Ni)=\int_0^{2\pi}\int_0^{\pi/2}\int_0^{\infty} \mathrm{KT}.\left(\frac{\mathrm{A}\cos\theta}{\rho_2}\right).e^{-\rho/\lambda'}.s.\left(\frac{d\sec\theta}{\lambda}\right).\rho^2\sin\theta\,d\rho\,d\theta\,d\phi$$

$$=2\pi\,\mathrm{KTA}s\,\lambda'\,d/\lambda.$$

If two members of a sheaf simultaneously produce a star in the emulsion, we have a pair of stars with a projected separation $\leqq \alpha\rho\sec\theta$. Let us take $\alpha\rho\sec\theta \leqq 2\,r_p$ as the condition defining a sheaf so that all stars produced in the emulsion by such a sheaf will have an average projected separation $\leqq r_p$. Then the number of double stars with projected separation $\leqq r_p$ is

$$\mathrm{R}=\int_0^{2\pi}\int_0^{\pi/2}\int_0^{2r_p/\alpha\sec\theta} \mathrm{KT}.\left(\frac{\mathrm{A}\cos\theta}{\rho^2}\right).e^{-\rho/\lambda'}\frac{s(s-1)}{2}\left(\frac{d\sec\theta}{\lambda}\right)^2.\rho^2\sin\theta\,d\rho\,d\theta\,d\phi$$

$$=2\pi\,\mathrm{KTA}.\frac{s(s-1)}{2}\frac{\lambda' d^2}{\lambda^2}\int_0^{\pi/2}\left[1-\exp-\left(\frac{2r_p\cos\theta}{\alpha\lambda'}\right)\right]\tan\theta\,d\theta$$

$$\mathrm{R}/(\Sigma Ni)=\frac{(s-1)}{2}.\frac{d}{\lambda}.\int_0^{\pi/2}\left[1-\exp-\left(\frac{2r_p\cos\theta}{\alpha\lambda'}\right)\right]\tan\theta\,d\rho$$

$$\equiv\frac{(s-1)}{2}.\frac{d}{\lambda}.f(r_p)\equiv \mathrm{C}.f(r_p)$$

The integral can be evaluated graphically if we assign a value for $\alpha\lambda'$. Fig. 2 shows the curves of $\mathrm{C}.f(r_p)$ against r_p for $\alpha\lambda'=(1/2000)$ cm., (1/200) cm., and (1/20) cm. respectively. They are normalized so that $\mathrm{C}.f(r_p)=21/1000$ at $r_p=(1/10)$ in. The factor C is then 0·0027, 0·0040, and 0·0075 respectively. The small circles represent the experimental values of $(\mathrm{P}-\mathrm{Q})/(\Sigma Ni)$. It can be seen that the shape of the curve changes very slowly with the values of $\alpha\lambda'$. It is not a sensitive method to fix the value of $\alpha\lambda'$ by finding the curve which fits the experimental data best.

If we assume $\lambda=18$ cm. (corresponding to a cross-section of 10^{-24} cm.^{2} for the average nucleus in the emulsion), s can be calculated from the value of C. Thus we have

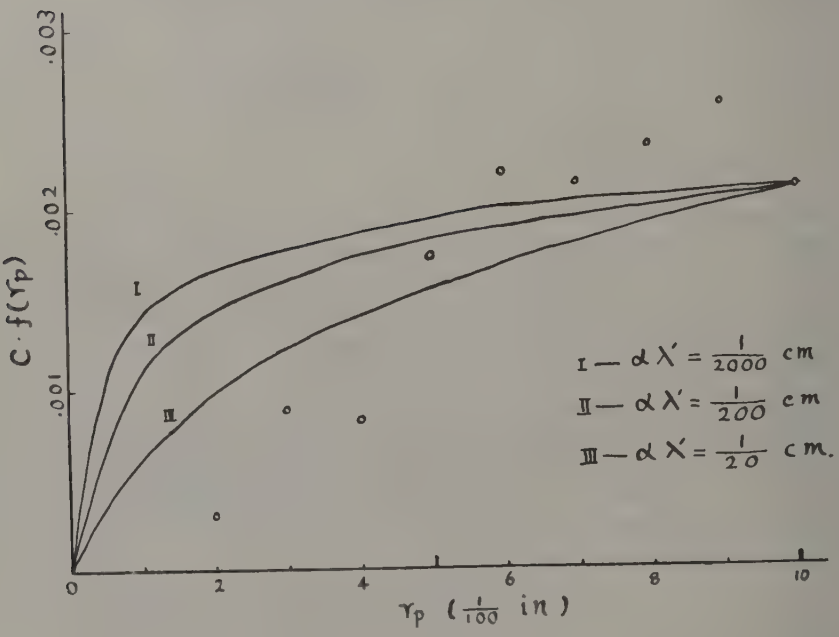
$\alpha\lambda'$	(1/20) cm.	(1/200) cm.	(1/2000) cm.
C	·0075	·0040	·0027
s	28	15	11

We see that if we want to make s reasonably small, $\alpha\lambda'$ would have to be extremely small. It is very difficult, indeed, to explain the double stars by this process even if we assume that all stars are produced in this way. But suppose we accept that $s=15$ and $\lambda=18$ cm. are reasonable, then we can explain $\alpha\lambda'=(1/200)$cm. by assuming a small conical angle, say $\alpha=(1/1000)$ rad., giving $\lambda'=5$ cm. which means that the particles are unstable, having a very short life-time. Thus we have tentatively the following set of values: $\lambda=18$ cm., $\lambda'=5$ cm., $\alpha=(1/1000)$ rad., and $s=15$, which can roughly account for the observed number of double stars.

§4. Conclusion.

It was seen in the last section that it is very difficult to account for the observed number of double stars with our present knowledge. The isotropic regenerative process accounts for at most 15 per cent of the effect. Collimated penetrating showers, especially if in cascades, may also account for a fraction of the effect, though it could only be a small one owing to their large conical angles. It is, however, possible that these double stars are due to something hitherto unknown. For instance, neutral particles emitted in large number from primary stars producing secondary stars, or production of stars by sheafs of unstable neutral particles with small conical angles such as appear to produce the close pairs of electrons recently observed (Carlson *et al.* 1950).

Fig. 2.



Acknowledgment.

The author wishes to express his sincere thanks to Prof. Sir George Thomson for his encouragement and help in the completion of this paper.

References

Brown, Camerini, Fowler, Heitler, King and Powell, 1949, *Phil. Mag.*, **40,** 862.
Carlson, Hooper and King, 1950, *Phil. Mag*, **41,** 701.
Leprince-Ringuet and Heidmann, 1948, *Nature, Lond.*, **161,** 844.
Li and Perkins, 1948, *Nature, Lond.*, **161,** 844.

CIII. *Communication Theory and Physics.*

By D. GABOR,
Imperial College, London*.

[Received August 10, 1950.]

SUMMARY.

The electromagnetic signals used in communication are subject to the general laws of radiation. One obtains a complete representation of a signal by dividing the time-frequency plane into cells of unit area and associating with every cell a " ladder " of distinguishable steps in signal intensity. The steps are determined by Einstein's law of energy fluctuation, involving both waves and photons.

This representation, however, gives only one datum per cell, *viz.*, the energy, while in the classical description one has two data : an amplitude and a phase. It is shown in the second part of the paper that both descriptions are practically equivalent in the long-wave region, or for strong signals, as they contain approximately the same number of independent, distinguishable data, but the classical description is always a little less complete than the quantum description. In the best possible experimental analysis by an electronic device the number of distinguishable steps in the measurement of amplitude and phase is only the fourth root of the number of photons. Thus it takes a hundred million photons per cell in order to define amplitude and phase to one per cent each.

COMMUNICATION theory has up to now developed mainly on mathematical lines, taking for granted the physical significance of the quantities which figure in its formalism. But communication is the transmission of physical effects from one system to another, hence communication theory should be considered as a branch of physics. Thus, it is necessary to embody in its foundations such fundamental physical data as the quantum of action, and the discreteness of electric charges. This is not only of theoretical interest. With the progress of electrical communications towards higher and higher frequencies, we are approaching a region in which quantum effects become all-important. Nor must one forget that vision, one of the most important paths of communication, is based essentially on quantum phenomena.

Some years ago, I have proposed a mathematical framework for the representation of signals (Gabor 1946). I have been rightly criticized for having left out noise, which is an essential feature of all communications. This will be remedied here, and at the same time the description will be brought in line with modern physics. But as the mathematical frame will serve as a useful foundation, it will be necessary to give first a short review of it.

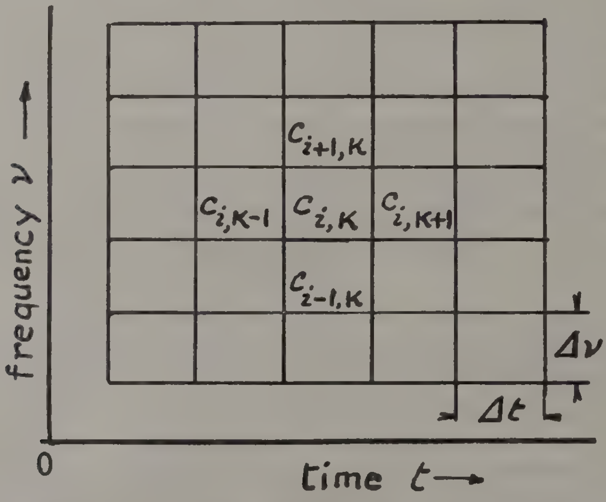
---

* Communicated by the Author.

§1. CLASSICAL REPRESENTATION OF SIGNALS.

The previous work started from the observation that the description of a signal in the conventional way, as a continuous function of time, is redundant and non-physical. A continuous function contains in any interval, however small, an infinity of data, corresponding to an infinite range of frequencies. A similar objection can be raised against the Fourier representation, which involves infinite time. In the new description one considers the signal simultaneously as a function of frequency *and* time. It is convenient to use only positive frequencies in the description. This can be done by introducing a certain complex function, whose real part is the physical signal. (The theory of these complex or " analytical " signals has in the meantime received interesting additions by the work of J. A. Ville (1948)). If the time-frequency half-plane, fig. 1, is divided, by any network, into cells of unit area, $\Delta t \Delta \nu = 1$, one finds that the signal in any domain containing a sufficiently large number of cells is fully described by associating two real data, or one complex datum, with every cell. In other words, each cell has two degrees of freedom.

Fig. 1.

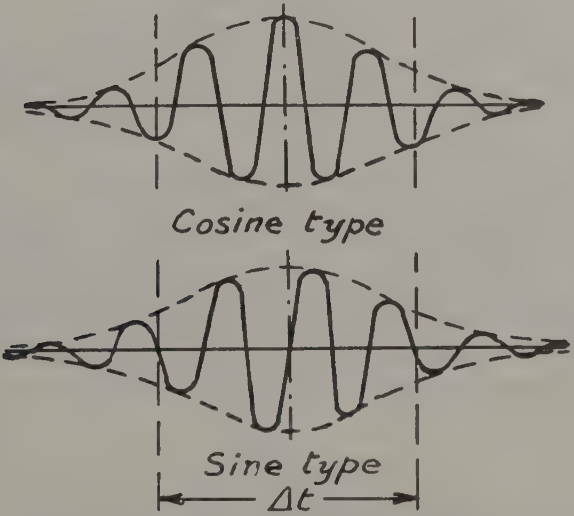


Information diagram. The time–frequency half-plane is divided up into cells of unit area and an elementary signal is associated with each, with a coefficient c_{ik}.

One can represent an arbitrary signal in an infinity of ways as a linear function of certain " elementary " signals, associated with the individual cells. There is, however, one description of particular interest, in which the elementary signals are harmonic functions, modulated with a " gaussian " signal, *i.e.*, they have envelopes of probability shape.

(Fig. 2.) These share with other functions the property that their Fourier transforms have the same shape, but they are unique in that respect that the product of their " effective " duration and of their effective frequency width is the smallest possible for any function. Thus, it can be said that these elementary functions overlap as little as possible. They have also the advantage that the familiar concepts " amplitude " and "phase" can be used in connection with them in the same way as with infinite harmonic functions. The complex elementary function is Cos $+_j$Sin, if we denote by Cos the even, and by Sin the odd type of real elementary signal. Multiplying these with suitable complex coefficients c_{ik}, as indicated in fig. 1, we can represent any arbitrary signal. Time-description and frequency-description (Fourier integral) can be considered as extreme special cases of this representation. In the first case the even elementary signal degenerates to a delta-function, and the odd one to its derivative, in the second case both become ordinary harmonic functions.

Fig. 2.



Elementary signals of the " cosine type " (even), and of the " sine type " (odd).

The " matrix " representation, illustrated in fig. 1, is proof against the objections raised against the pure " time " or " frequency " descriptions, but it does not go far enough. The infinity of data has been reduced to a finite number, but these data, *i.e.*, the coefficients c_{ik}, are still supposed to be *exactly* defined. But a single exact datum still contains an infinite amount of information, *i.e.* an infinite number of " yeses or noes " In reality, of course, these amplitudes, like every physical datum, have a certain amount of uncertainty or " noise." This has been taken into consideration in the mathematical theories of Shannon (1948) and Tuller (1949), where the noise amplitudes are assumed as known. But we cannot be satisfied with this in a physical theory. Even if all accidental imperfections of the instruments are eliminated, there remain certain basic uncertainties, which we are now going to investigate.

§2. STATISTICAL PROPERTIES OF THE INFORMATION CELL.

In order to connect the mathematical scheme with physical reality, we must first observe that every physical signal has a certain energy associated with it. We can also associate a certain energy with every cell *, and, for simplicity, we will say that it is "contained" in the cell.

We will limit the discussion to electric communications, though most of our results will apply equally well to sound, or, in short, to communication by any quantity which is considered as continuous in classical physics. We observe first that all electric signals are conveyed by *radiation*. Even if lines or cables are used in the transmission, by the Maxwell-Poynting theory, the energy can be located in empty space. Hence we can apply to our problem the well-known results of the theory of radiation.

For simplicity, we consider our communication system as having the uniform temperature T. The uncertainty connected with the concept of temperature will produce certain fluctuations in the energy of the cells, which we can calculate by the rules of statistical thermodynamics once we know the law for the mean thermal energy $\bar{\epsilon}_T$ of a cell, in function of the temperature. This we obtain at once, if we observe that every cell in the signal has two degrees of freedom, of which only one *counts* for the purpose of statistics, as the other is of the nature of a phase. Thus, by Planck's law

$$\bar{\epsilon}_T = \frac{h\nu}{\exp(h\nu/kT) - 1}, \qquad (1)$$

where ν is the mean frequency of the cell. In other words, we identify every information cell with a "Planck oscillator".

This requires perhaps a little more explanation, as physicists are less familiar with a discussion of radiation in terms of frequency and time than in terms of frequency and space. But the first case is immediately reduced to the second if we imagine the signal propagating with the velocity c, and plot the information diagram against the length ct instead of against the time t. The state of the field in such a linear system

---

* The energies of the elementary signals of the type discussed will not add up exactly to the energy of the whole signal, because they are not quite orthogonal, but the error will vanish in the mean over large number of cells. One can, however, to meet objections, consider instead an orthogonal set of elementary signals, such as the signals with "limited spectrum", introduced by Shannon (1949) and by Oswald (1949, 1950) which have uniform spectral density inside and zero outside a frequency strip. This makes no difference to the following discussion, as no reference will be made to any special type of elementary signal.

(in which we consider one state of polarization only) can be represented by two systems of progressive waves in opposite directions, only one of which represents the signal in which we are interested. We count the degrees of freedom in this wave system—which is what we have done in the last section—and give it the energy $\bar{\epsilon}_{T}$ for each amplitude, disregarding the phases. This is the application to the linear case of von Laue's well-known derivation of Planck's law for the radiation density by superposition of plane waves (von Laue 1914).

The thermal energy $\bar{\epsilon}_{T}$ does not in itself represent "noise", *i.e.*, an uncertain disturbing factor. It becomes disturbing only by its fluctuations. In order to obtain the mean square fluctuations of the energy, we apply Einstein's formula

$$\overline{\delta\epsilon_{T}^{2}}=\overline{(\epsilon_{T}-\bar{\epsilon}_{T})^{2}}=kT^{2}\frac{d\bar{\epsilon}}{dT}T, \qquad (2)$$

which gives

$$\overline{\delta\epsilon_{T}^{2}}=h\nu\bar{\epsilon}_{T}^{2}+\bar{\epsilon}_{T}. \qquad (3)$$

Einstein's interpretation of this equation is well known (*cf.* Born 1949). The second term has been identified by H. A. Lorentz with the fluctuations due to the interference of waves with random phases.* But the first term suggests that the energy is concentrated in light quanta or photons of energy $h\nu$, which fluctuate in any element as if they were independent particles. As both effects are present simultaneously, and are acting as if they were independent of one another, it is not possible to make a simple physical picture of the process. But fortunately this is not necessary. Applying Einstein's argument to the case when a signal is present, so that the mean energy $\bar{\epsilon}$ in the cell exceeds the thermal mean energy $\bar{\epsilon}_{T}$, we obtain in Appendix I.

$$\overline{\delta\epsilon^{2}}=\overline{(\epsilon-\bar{\epsilon})^{2}}=h\nu\bar{\epsilon}+2\bar{\epsilon}\epsilon_{T}-\bar{\epsilon}_{T}. \qquad (4)$$

Expressing the energy by the number of photons N in the cell, so that $\epsilon=Nh\nu$, $\bar{\epsilon}_{T}=\bar{N}_{T}h\nu$, this becomes

$$\overline{\delta N^{2}}=\overline{(N-\bar{N})^{2}}=\bar{N}(1+2\bar{N}_{T})-\bar{N}_{T}^{2} \qquad (5)$$

with

$$\bar{N}_{T}=1/[\exp(h\nu/kT)-1]. \qquad (6)$$

---

* In technical theories of thermal noise, it is usually forgotten that it is not the noise power but its fluctuations which cause the disturbance. But if the quantum effect is small and the second term in equation (3) predominates, the r.m.s. value of the noise power fluctuation is equal to the noise power itself, hence this error is without consequences.

From equation (3), neglecting the quantum term, one can easily derive Nyquists well-known rule (Nyquist 1928) that a resistance R can be considered as containing a " noise generator " with a mean square electromotive force

$$\overline{E^{2}}=4kTR\Delta\nu.$$

The proof can be given in the same form as Nyquist has done, by substituting a cable with wave impedance R for the resistance. But one must add the condition that not only the noise power, but also its fluctuations follow the same rule in the resistance as in the cable. This is no arbitrary rule; the necessity of a universal law of mean square fluctuations follows directly from the second principle, as Szilárd (1925) has shown.

$\overline{N}_T$, the number of " thermal " photons per cell, is a large number for the frequencies used in electrical communications. One can use the approximate formula

$$\overline{N}_T=\frac{kT}{h\nu}=\frac{kT}{hc}\lambda=0{\cdot}7\lambda T$$

which gives $\overline{N}_T=210$ for a wavelength of 1 cm. and a temperature of 300° K. On the other hand for visible light, $\lambda=5{\cdot}10^{-5}$ cm., the approximate formula

$$\overline{N}_T=\exp(-h\nu/kT)=\exp(-1/0{\cdot}7\lambda T)$$

gives $\overline{N}_T=e^{-94}=10^{-39}$. Hence for visible light, at ordinary temperatures, there is practically no thermal noise, and the fluctuation becomes pure " quantum noise ", which follows the law $\overline{\delta N^2}=\overline{N}$.

These results enable us to construct a complete physical representation of a signal. We see that the state of an information cell is completely determined by a stochastic number N, the number of photons. We can now construct a scale or "*ladder*" *of distinguishable states*, on which every step corresponds to a reasonably ascertainable difference. It is an evident suggestion to adopt the r.m.s. fluctuation of N as the unit step.* With this convention, the number of steps distinguishable below a maximum level N_m is, by equation (5),

$$S=\int_{\overline{N}_T}^{N_m}\frac{dN}{(\overline{\delta N^2})^{1/2}}=\int_{\overline{N}_T}^{N_m}\frac{dN}{[N(1+2\overline{N}_T)-\overline{N}_T^2]^{1/2}}$$

$$=\frac{2}{1+2\overline{N}_T}\{[N_m(1+2\overline{N}_T)-\overline{N}_T^2]^{1/2}-[\overline{N}_T(1+\overline{N}_T)]^{1/2}\}\approx\frac{2N_m^{1/2}}{(1+2\overline{N}_T)^{1/2}}. \quad . \quad (7)$$

The last formula is valid for large signals.

In the useful terminology introduced by D. M. MacKay (1950), S is the " proper scale " of the photons. Fig. 3 illustrates the representation of a signal in three dimensions, with the photon scale at right angles to the time–frequency plane. If instead we plotted $\log_2 S$, the ordinates would give directly the equivalent number of binary selections or " bits " This is the number of "yeses" or "noes" required to fix the position of the signal on the ladder.

It is of interest to enquire about the minimum energy required for the transmission of the first "bit" of information. By our convention, this cannot be less than the r.m.s. value of the thermal energy fluctuations, though it can be more. Combining equations (1) and (3), one obtains

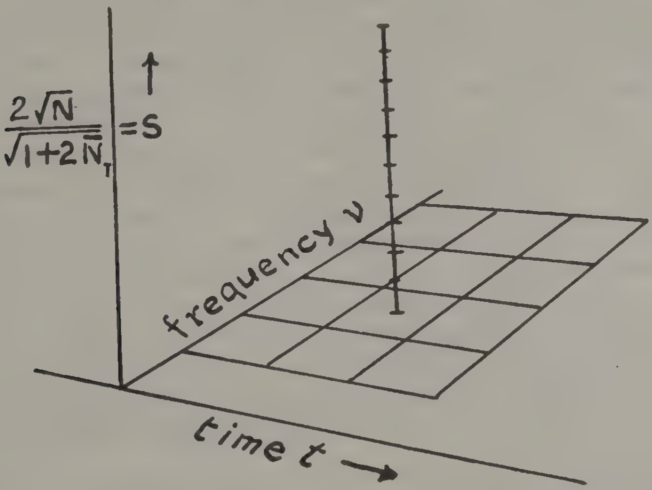
$$\epsilon_{\min}=(\overline{\delta\epsilon_T^2})^{1/2}=\frac{h\nu}{2\sinh(h\nu/2kT)}. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

---

* By the theorem of Bienaymé and Tchebycheff, the probability of an error k times exceeding the r.m.s. error is smaller than $1/k^2$, whatever the law of the fluctuations may be.

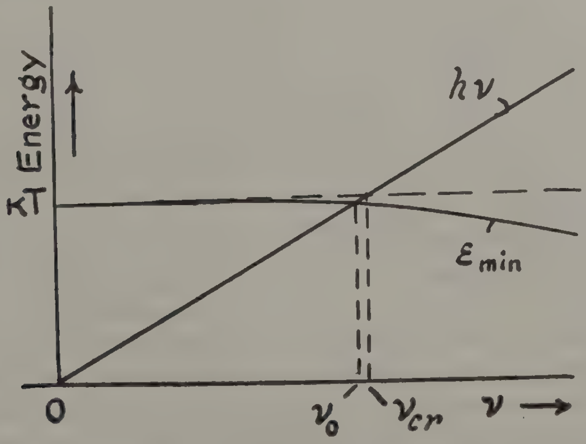
But, on the other hand, this energy cannot be less than one quantum, $h\nu$. Thus the energy required for the first " bit " is either $\epsilon_{\min}$ as given by equation (8), or $h\nu$, whichever is the larger of the two. As shown in fig. 4, the two lines cross at $h\nu_0=0{\cdot}96kT$, which is only slightly less than kT.

Fig. 3.



Representation of an arbitrary signal. A ladder or " proper scale " of distinguishable values is erected in every cell, on which the occupation is marked off.

Fig. 4.



The energy required for the transmission of the first " bit " of information.

Thus up to a critical frequency ν_{cr}, an energy kT is sufficient for the first step, but no communication is possible with an energy of less than kT.

The interesting feature of this result is its generality; it applies to the unknown processes in the nervous system as well as to electrical communications.

§3. STATISTICS OF SIGNALS.

Up to this point, we have directed our attention only to one cell. A short digression on signals and ensembles of signals may not be out of place before we return to the physical analysis of our results.

A signal is a system of, say, n cells, preferably, but not necessarily, contiguous in the information plane. We consider a large number of such systems in a stationary transmission, in which they differ only by their position in time, not in frequency, and we speak of them as of an "ensemble". Evidently the analogy with statistical mechanics is not very perfect. In statistical mechanics, we can either follow a system in an ensemble over a long time, or we can look at all systems in the ensemble simultaneously, while here we can take only the first view. It is also somewhat questionable whether the often used expression "ergodic" is justified. In its original sense, due to Willard Gibbs, it means that each system (each signal) spends equal times in all states compatible with a given energy, which is not true for most stationary transmissions usually considered as "ergodic" in communication theory. Thus, we prefer to avoid this term.

The most important mean value in such an ensemble is that of the *entropy*. In order to clarify the connexions between physics and communication theory, it may be useful to consider this problem in two stages. In the first, we consider all configurations of the system of n cells, compatible with the energetic conditions of the transmission as equally probable, and define the entropy as $k \log P$, where P is the number of all these possible configurations. In the second stage, however, we give them different probabilities or "weights". The first stage is in close connection with physics; actually it is the problem of calculating the entropy of an "ergodic" system, in the original meaning of the word. According to quantum statistics, all simple, accessible states have equal probability, and the levels of Planck oscillators are, of course, simple. (*Cf.* Jordan 1933.)

As an example, let us estimate the mean entropy of groups of n cells in a transmission in which the mean energy level is $\overline{S^2}$, and the r.m.s. deviation from the mean is ΔS_n, for brevity. (The suffix n had to be added, as the standard deviation is dependent on the size of the group.) For simplicity, we assume $(\Delta S_n)^2 \ll \overline{S^2}$, *i.e.*, a small degree of "modulation". This allows us to equate the number P of possible states to the number of points with positive, integer coordinates inside a spherical shell in n-dimensions, whose mean radius R is given by

$$R^2 = n\overline{S^2},$$

while its thickness is

$$2\Delta R = n\Delta S_n/R.$$

Well-known formulæ for the volume of a n-dimensional sphere give

$$\mathrm{P}=\frac{(\pi/4)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n+1)}\,n\mathrm{R}^{n-1}2\Delta\mathrm{R}$$

$$=\frac{(\pi/4)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n+1)}\,n(n\overline{\mathrm{S}^2})^{1/2n}\,\frac{\varDelta\mathrm{S}_n}{\overline{\mathrm{S}^2}}\approx\frac{1}{\sqrt{\pi}}\,(\tfrac{1}{2}e\pi\overline{\mathrm{S}^2})^{\frac{1}{2}n}\,\frac{\sqrt{(n)}\,\varDelta\mathrm{S}_n}{\overline{\mathrm{S}^2}}. \quad . \quad (9)$$

We have used Stirling's formula to approximate the Γ-function for large arguments. For sufficiently large n, however complicated the signals, they must behave as if they were independent, and we must have asymptotically

$$n\,.\,(\varDelta\mathrm{S}_n)^2\rightarrow\mathrm{const.}\,;$$

hence, apart from a constant, the entropy *per cell* goes to

$$k\log\,(\tfrac{1}{2}e\pi\overline{\mathrm{S}^2})^{1/2}, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (10)$$

a result also obtained by Shannon (1949).

Apart from the factor k, the entropy (10) is also a measure of the information capacity per cell of the transmission, in accordance with its definition as the logarithm of the number of possible, equally probable, selections. Thus, the concepts of information capacity and of entropy are closely parallel, even identical.

Consider now for a moment the central problem of statistical communication theory, which has been investigated by Shannon, the communication of symbols, as supplied by some source. If the source supplied them with equal probability, one could connect the possible and equally possible states of the signal with the symbols by an infinity of possible codes, and transmit information at the maximum rate, given by (10) multiplied by the number of cells per second. But if, as it is mostly the case, the symbols have different probabilities p_i, it would be uneconomical to give them equal shares in the available signal entropy. Those which occur more often, and hence carry less information, should be given less than the more unexpected ones. In this way, though it is not, of course, possible to make the channel carry more information than given by (10), it is possible to make it carry *more symbols*, each symbol in the average being worth less. Shannon (1948) has shown that in the case of optimum coding each bit of channel capacity can carry 1/H symbols, where H is defined as

$$\mathrm{H}=-\Sigma p_i\log p_i\ \text{bits/symbol}, \quad . \; . \; . \; . \; . \; . \quad (11)$$

i.e., as the mean value of a symbol in terms of binary selections. This "*selective entropy*" is defined by the source of symbols, and it is at an appreciable remove from the physical entropy of the signal previously discussed, as may be already seen from the fact that it appears as its divisor. The interesting properties of the expression (11), demonstrated by Shannon, follow rather from its mathematical form than from its intrinsic relation with Boltzmann's "H".

There exists, however, a general limiting relation between selective information and entropy in the thermodynamical entropy of the source considered as a physical system, which was revealed by Szilárd (1929). This is that a signal corresponding in information value to an s-fold selection enables the receiving system to reduce the entropy of the transmitter by a *maximum* of $k \log s$. It follows, as he proved in detail, that any mechanism acquiring this information must increase the entropy by a *minimum* of $k \log s$, thus safeguarding the second principle of thermodynamics.

§ 4. THE CONNECTION BETWEEN QUANTUM REPRESENTATION AND CLASSICAL DESCRIPTION.

Comparing the representation of a signal, illustrated in fig. 3, with the classical, mathematical description, it appears that we have lost something. Previously we had two data per cell; we are left now with one only, a function of the quantum number N, which corresponds to the energy. What has happened to the phase ?

It is not difficult to give an answer to this question on general lines. N is not an exact datum but a stochastic number, which represents a finite amount of information. But between one datum of finite accuracy and two, there is no transcendental abyss of the kind which exists between a simple and a twofold infinity. One finite datum can very well contain two independent, finite data, provided that their aggregate information is not more than the original. It will be shown that this is indeed the case, and that the maximum amount of information on amplitude and phase are both contained in the single photon scale.

An electromagnetic signal can be physically analysed in various ways, but it will be instructive to consider first extreme cases only. One extreme is a *counter*, an instrument which records single photons. It follows immediately from Heisenberg's uncertainty principle that the "time resolution" of such an instrument cannot be better than a whole cycle, hence the phase remains entirely unobservable. The other extreme is any classical field-measuring instrument, capable of recording the electromagnetic field in the signal as a function of time. It may be called for brevity a proportional *amplifier*, as amplification of weak signals is one of its essential functions. There is no need to consider intermediate instruments, as it will be seen that in a certain range of very weak signals every classical amplifier operates as a sort of "proportional counter".

One limit for the operation of any such instrument is given by the well-known uncertainty relation of the quantum theory of radiation (*cf.* Heitler 1944, p. 68)

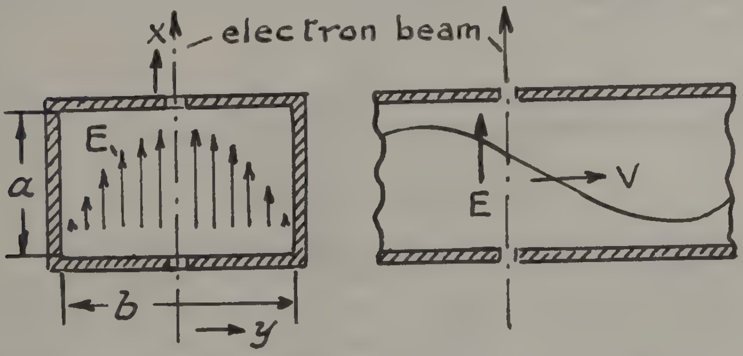
$$\Delta N \Delta\phi \geqslant 1, \qquad (12)$$

where ΔN is the uncertainty in the photon number N, and $\Delta\phi$ the uncertainty in the phase ϕ. But this gives merely an upper limit in our case, as we want to determine amplitude and phase *simultaneously*, each as accurately as possible. It will be seen that, in fact, the limiting accuracy which is practically obtainable is much below what might be expected from the inequality (12).

We will approach the problem by a detailed analysis of a particular type of proportional detector-amplifier. Subsequently, we will try to improve its performance to the extreme limit. It will then be found that all special features of the device vanish from the formulæ, thus the final result can be considered as of general validity.

Assume that the electromagnetic signal is introduced into a rectangular wave guide, in the TE_{10} mode, *i.e.*, with an electric field in the x-direction, independent of the x-coordinate, which is at right angles to the direction of propagation z. An electron beam passes through two small holes along the x-axis, in the plane of maximum intensity, with velocity v and current intensity J. The alternating acceleration and retardation of the electrons produces velocity modulation in the beam, which by well-known methods can be translated into current modulation, *i.e.*, producing an alternating component superimposed on the mean value J. One method utilizes the " bunching " of the electrons which takes place at a certain distance from the guide, but one can also deflect the beam by a constant field and make it play between two collecting electrodes close together. In either case, one obtains an alternating current which in first approximation is proportional to the d.c. current and to the relative accelerations and retardations suffered by the electrons in the wave guide.

Fig. 5.



Analysis of an electromagnetic signal in a wave guide by an electron beam.

Let us measure the energy exchange between the electrons and the field in quantum units, $h\nu$, where ν is the mean frequency of the signal. (It will soon be seen that, in order to make the exchange intense, the waveband of the signal must be made so narrow that it is permissible to take the arithmetic mean.) Let $\overline{N}$ be the mean number of photons in an information cell, which passes through a cross-section of a guide in a time $\Delta t = 1/\nu$. During this time $\overline{N} = J\Delta t/e$ electrons pass through it, and exchange in the mean $\overline{n}$ quanta with the field, either by losing or by gaining energy. The positive number $\overline{n}$ is the essential parameter of the process. If $\overline{n}$ is a large number, which is possible only if $\overline{N}$ is also large, the interchange will be essentially classic; if $\overline{n}$ is small, quantum phenomena will dominate.*

The detailed calculations may be found in the Appendices II. and III., only the results will be discussed here. The first contains the classical theory of exchange, valid for large $\overline{n}$ and $\overline{N}$, the second a wave-mechanical calculation valid for small exchange parameters. In the classical theory the result is

$$\overline{n}^2 = \frac{32}{\pi^3}\left(\frac{2\pi e^2}{hc}\right)\frac{\sin^2\theta}{\theta}\,\frac{v}{\nu b}\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{-1/2}\frac{\Delta\nu}{\nu}\,\overline{N}. \quad . \quad . \quad (13)$$

Thus the exchange parameter is proportional to the square root of the photon number, as may be expected. Of the dimensions of the wave guide the width b appears explicitly in the factor of $\overline{N}$, while the depth a is contained in the angle

$$\theta = \pi\nu a/v \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

which is one-half of the "transit angle". c is the velocity of light. In addition, there appears a factor which is the reciprocal of

$$\frac{hc}{2\pi e^2} = 137,$$

the fundamental number which connects photons with electrons.

Anticipating that the best measurement will require an intense interchange, we now try to increase the coefficient of $\overline{N}$ in equation (13) by all available means. First we make the factor $\sin^2\theta/\theta$ a maximum. This is 0·723 and is obtained with $\theta = 67°$, *i.e.*, a transit angle of 134°. This disposes of the depth a of the wave guide. The optimum width b is determined by the condition that the group velocity

$$U = c\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}$$

must be as small as possible. But the smallest value is reached when U vanishes at the low-frequency limit, $\nu - \frac{1}{2}\Delta\nu$, of the band. Substituting those values into equation (14), one obtains in the optimum case

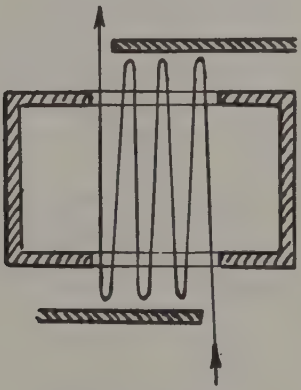
$$\overline{n}^2 = \frac{1{\cdot}5}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}}\overline{N}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

---

* Quantal energy exchange between electrons and the field in a wave guide at high quantum numbers has been previously discussed by Lloyd P. Smith (1946), but we cannot agree with most of his results. Monokinetic electrons and exchange of sharply defined quanta on the one hand, well-defined entrance phases and short transit times on the other, are mutually exclusive phenomena by the Uncertainty Principle; hence we believe that only certain averages over Smith's detailed results have physical significance.

All special features of the device have vanished in this formula, apart, perhaps, from the unimportant factor, 1·5. But it is evident that the factor of $\overline{N}$ must always be much smaller than unity, while its best value as will be shown later, is just unity. There exists, however, a further possibility for improving the performance. Assume that we can make each electron perform repeated passages through the guide, each transit, in the opposite direction exactly half a cycle after the last. (This is possible in principle, as the optimum transit angle is about 134°.) If the frequency were known beforehand, the number of passages would be limited only by the consideration that, by repeated gains or losses, the electron would be bound to fall out of synchronism. But if the signal had a single frequency, known in advance, there would of course be no communication. However, even if the frequency is known beforehand only within $\nu \pm \frac{1}{2}\Delta\nu$, one can make the number P of passages as great as $\nu/\Delta\nu$, without risking a phase error of more than $\pm\frac{1}{2}\pi$, and it can be shown that these passages are almost of equal value, so that $\bar{n}$ is increased very nearly by a factor P. The number of passages required to make the coefficient of $\overline{N}$ in equation (15) unity is

Fig. 6.



Repeated passages.

$$P_{\mathrm{opt}} = 9{\cdot}5\left(\frac{v}{c}\right)^{1/2}\left(\frac{\nu}{\Delta\nu}\right)^{1/4}, \quad \ldots\ldots\ldots \quad (16)$$

i.e., at least of the order ten. But this number must not exceed $\nu/\Delta\nu$; hence we obtain the condition that, in order to realize optimum conditions, the frequency band must be so restricted that

$$\left(\frac{\nu}{\Delta\nu}\right) > 20{\cdot}2\left(\frac{c}{v}\right)^{2/3}. \quad \ldots\ldots\ldots \quad (17)$$

This is a somewhat surprising result. In the mathematical representation, it did not matter whether we divided up the frequency band into broad or narrow strips. But by the intervention of the number 137, ($20{\cdot}2 = (137/1{\cdot}5)^{2/3}$), it turns out that only narrow frequency bands are capable of accurate analysis by means of electrons. (If ions with charge Z were used, one would have to replace 137 by $137/Z^2$, and the condition would be less stringent.)

Thus, at least in theory, the device could be perfected up to the optimum performance. The practical difficulties are of course evident. In practice one would rather replace the wave guide by a "high-Q" resonator, but this would somewhat complicate the theory.

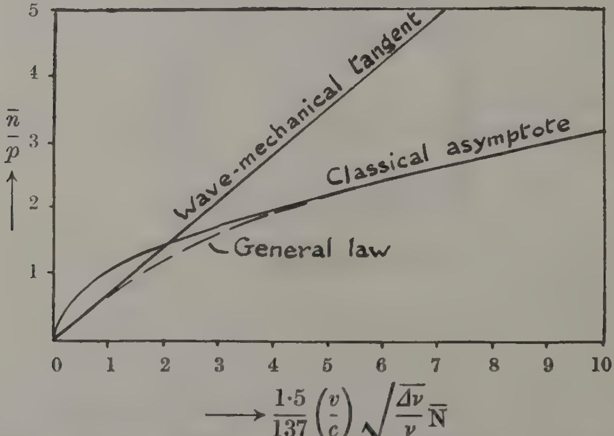
We thus find, assuming P passages, in the "classical" case

$$\left(\frac{\bar{n}}{\mathrm{P}}\right)^2 = \frac{1\cdot5}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}}\overline{\mathrm{N}}, \qquad (18)$$

while the corresponding wave-mechanical formula, valid for very weak interchange, is (Appendix III.)

$$\frac{\bar{n}}{\mathrm{P}} = \frac{1\cdot0}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}}\overline{\mathrm{N}}. \qquad (19)$$

Fig. 7.



Mean quantum exchange between electrons and photons. The initial tangent is calculated from wave mechanics, the asymptotic parabola classically.

Apart from the factor 2/3, the coefficient of $\overline{\mathrm{N}}$ is the same in both cases, but this time the quantum exchange $\bar{n}$ is proportional to the photon number itself, not to its square root It can be said, therefore, that *for small photon concentrations, the device acts as a counter, at large concentrations, as a field-measuring instrument.* The intermediate region is difficult to calculate, but, as shown in fig 7, the two branches can be connected by a plausible curve.

One can also interpret the results in this way: If there are few photons present, there will be few collisions, and—as shown in Appendix III. equal probabilities of gains and losses, at any instant. With increasing photon concentration, repeated collisions will increase in number, and the resulting loss or gain increases with the square root of the photon number only, but this resultant has now a prevailing direction, which changes its sign with the frequency of the signal. At this stage the " classical field " has developed.

Having ascertained that, with certain reservations, we can make the exchange as strong as we like, we ask the question: If we know the average photon number $\overline{N}$, how must we adjust the electron beam current J, and the exchange parameter $\bar{n}$ in order to measure the field amplitude E with maximum accuracy ? And having made these adjustments, how many steps shall we be able to distinguish in the scale of the field amplitudes ? Evidently this question has a precise meaning only in the " classical " range of large $\bar{n}$ and $\overline{N}$, and the following considerations relate only to this case. In order to simplify the problem, we neglect the thermal noise, *i.e.*, we put $\overline{N}_T=0$, so that the relative accuracy on the photon scale would be $1/\overline{N}^{1/2}$, and the total number of steps in the photon ladder $2\overline{N}^{1/2}$. The calculations are carried out in Appendix IV., here we give only the physical considerations

The quantity to be measured is the electric amplitude in the information cell, which is proportional to the square root of the photon number. The measured quantity, on the other hand, is the alternating electron current, which, as mentioned above, is proportional to $\bar{n}.J$ or to $n.\overline{M}$ for not too strong signals, $\overline{M}$ being the mean number of exploring electrons per cell. For the optimum, we impose the condition that the relative mean square deviation of the quantity measured from the quantity to be measured shall be as small as possible, *i.e.*,

$$\frac{\overline{(nM-CN^{1/2})^2}}{(\overline{nM})^2}=\min, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

where the proportionality factor C is determined from the condition

$$\overline{(nM-CN^{1/2})}=0.$$

We have to choose $\bar{n}$ and $\overline{M}$ so as to satisfy the condition (20) for given $\overline{N}$. There must be an optimum, for these reasons: A too weak interchange $\bar{n}$ would leave the cell unexplored. A too strong interchange, on the other hand, will *interfere with the object of the measurement* and spoil it. Though in the mean electrons are as often accelerated as retarded, fluctuations in the numbers M_1 and M_2 of electrons which pass through accelerating and retarding phases might produce extra photons, which could not be distinguished from those belonging to the signal, or annihilate some. The spurious photons are generated according to a law

$$\delta_2N=\bar{n}(M_1-M_2)=\bar{n}(\delta M_1-\delta M_2) \quad . \quad . \quad . \quad . \quad . \quad (21)$$

as $\overline{M}_1=\overline{M}_2=\frac{1}{2}\overline{M}$. We have written $\bar{\delta}_2N$ for this number, considering it as a fluctuation which must be added to the natural fluctuation δ_1N, whose law is $\overline{\delta_1N^2}=\overline{N}$. The two fluctuations must be considered as independent.

It is already evident from the above that the fluctuations in the number of beam electrons, *i.e.*, the " shot effect ", play an important part in these phenomena. A too weak current has a high relative fluctuation. A too strong current, especially aided by a large exchange, will again spoil the object. It may be noted that *we have here a type of uncertainty which springs directly from the fact that photons and electrons are discrete*, without any reference to the physical values of h and of e.*

The relative error according to equation (20) is calculated on the basis of equation (21), together with such evident assumptions as the independence of the fluctuations of n and M and of the "natural" part of δN. We assume also "normal" shot effect, $\delta M^2 = \bar{M}$. The result is

$$\frac{\overline{(nM - CN^{1/2})^2}}{(\overline{nM})^2} = \frac{1}{4\bar{N}} + \frac{1}{\bar{\bar{n}}} + \bar{n}\left(\frac{1}{\overline{nM}} + \frac{\overline{nM}}{4\bar{N}^2}\right). \quad . \quad . \quad . \quad (22)$$

This is a minimum for

$$\overline{nM} = \bar{n} . \bar{M} = 2\bar{N} \quad \bar{n}^2 = \bar{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

from which $\bar{M} = 2\bar{N}^{1/2}$. This gives the simple rule that, for optimum analysis of the signal, one must take one electron for every step in the scale of the photons, and the interchange $\bar{n}$ must be itself equal to one distinguishable step at the level $\bar{N}$. This again is a general rule, quite independent of the special model from which we started.

Substituting these values, we obtain for the minimum of the mean square relative error in the measurement of amplitudes

$$\left(\frac{\overline{(nM - CN^{1/2})^2}}{(\overline{nM})^2}\right)_{\min} = \frac{1}{4\bar{N}} + \frac{2}{\bar{\bar{n}}} = \frac{1}{4\bar{N}} + \frac{2}{\bar{N}^{1/2}}. \quad . \quad . \quad . \quad (24)$$

As we are dealing with large photon numbers only, the first term at the right-hand side can be neglected with respect to the second. Thus we see, applying the same rule which we have used in constructing the photon ladder, that the smallest distinguishable relative step in the amplitude scale is $\sqrt{2}$ times larger than the square root of the corresponding quantity in the photon scale. In other words, *the proper scale of the amplitudes will contain always less than the square root of the number of steps in the photon scale;* apart from the factor $\sqrt{2}$ for the reason that the optimum setting is, of course, possible for one level only.

Having determined the proper scale of the amplitudes, a simple application of the Uncertainty Principle (Appendix IV.) shows that the proper scale of the phase must also contain less steps than the square root of the photon scale. Thus, summing up, we see that the classical description of the signal, by being too detailed, gives in fact a somewhat *smaller total amount of information* than the quantum description.

The classical method of description, though theoretically inferior, may of course be still the best practically in the range of frequencies used for electrical communications, where efficient photon counters are not available. Conditions are different in the optical region, where detectors of the counter type—such as the eye—are not far from perfection. In this region, analysis in terms of electromagnetic waves is as yet technically impossible, but it is interesting to note that, even if it were possible, it would not be very practical. Progress in the field of microwaves is now actually approaching a region where the two different methods of analysis may become competitive. We see from our results that it takes about a hundred million photons per information cell in order to define amplitude and phase of the signal to 1 per cent each. Remembering that at 1 cm. wavelength the number of thermal photons per cell is only about 200, it may be seen that the time may not be far off when the imperfections of the classical method of description will manifest themselves even in electrical communications.

---

* It may be noted, however, that the shot effect, that is to say the random arrival of the exploring electrons, is a practical rather than a fundamental limitation. As Prof. R. E. Peierls has pointed out to me, it must be *ideally* possible to measure amplitude and phase simultaneously up to the limit given by the relation 12. In fact if one could make the exploring electrons arrive at regular intervals, all terms but the first on the right-hand side of equation 22 would vanish. But as it is practically impossible to suppress the shot effect to any significant degree, our results retain their practical importance.

Will it ever be possible to generate and to amplify light by an extension of microwave techniques ? Generation may not be unthinkable, but we can show at once that a light-amplifier for the reception of weak-light signals is impossible. It follows from our results that an electromagnetic signal in which amplitude and phase are defined with a relative accuracy ϵ must carry at least a power

$$h\nu^2\epsilon^{-4}(\Delta\nu/\nu). \qquad (25)$$

Applied to visible light with $\lambda=5000$ A., a line width of 1 A. and $\epsilon=0{\cdot}01$, we obtain 4·7 watts. Even if amplitude and phase were defined to 10 per cent only, the power would still be 0·0005 watts, a very high intensity in terms of visible light. This example demonstrates that *what we see* are always photons, not waves.

It may be hoped that these considerations have shown that the concepts of information theory may well prove their usefulness when applied to problems of physics.

REFERENCES.

BORN, M., 1949, *Natural Philosophy of Cause and Chance* (Oxford: University Press), p. 81.

CARSON, J. R., and FRY, T. C., 1937, *Bell Syst. T. J.*, **16**, 513.

GABOR, D., 1946, *Journ. I.E.E.*, **93**, III, 429; 1947, *Ibid.*, **94**, III, 369; 1947, *Nature, Lond.*, **159**, 591.

HEITLER, W., 1944, *The Quantum Theory of Radiation* (Oxford: University Press, 2nd edn.).

JORDAN, P., 1933, *Statistische Mechanik* (Braunschweig: Fr. Vieweg).

VON LAUE, M., 1914, *Ann. d. Phys.* [4] **44**, 1197.

MACDONALD, D. K. C., and KOMPFNER, R., 1949, *Proc. I.R.E.*, **37**, 1424.

MACKAY, D. M., 1950, *Phil. Mag.* [7] **41**, 189.

NYQUIST, H., 1928, *Phys. Rev.*, **32**, 753.

OSWALD, J., 1949, *C.R.*, **229**, 21.

OSWALD, J., 1950, *Cables et Transmission*, **2**, 197.

VAN DER POL, B., 1930, *Proc. I.R.E.*, **18**, 1194; 1946, *Journ. I.E.E.*, **93**, III, 153.

SHANNON, C. E., 1948, *Bell Syst. T.J.*, **27**, 379, 623; 1949, *Proc. I.R.E.*, **37**, 10.

SMITH, L. P., 1946, *Phys. Rev.*, **69**, 195.

SZILÁRD, L., 1925, *Zeitschr. f. Phys.*, **32**, 753, 1925; 1929, *Ibid.*, **53**, 840.

TULLER, W. G., 1949, *Proc. I.R.E.*, **37**, 468.

VILLE, J. A., 1948, *Cables et Transmissions*, **1**, 61.

APPENDIX I.

ENERGY FLUCTUATIONS IN AN INFORMATION CELL IN THE PRESENCE OF A SIGNAL.

As in the case of thermal equilibrium, the mean square energy fluctuation is the sum of a classical and of a quantum term.

The classical term. Consider, for simplicity, a (complex) Fourrier-component E_S of the signal, and a component E_T of the thermal noise, corresponding to two different (circular) frequencies, say, ω_1 and ω_2. The instantaneous energy density, resulting from the interference, is proportional to

$$E_S E_S^* + E_T E_T^* + \{E_S E_T^* \exp[j(\omega_1 - \omega_2)t] + E_S^* E_T \exp[-j(\omega_1 - \omega_2)t]\},$$

where the asterisks stand for the conjugate complex values. The first two terms are the energy of the signal and the energy of the noise, the rest arises from interference. (Beats.) Let us write

$$\epsilon = \epsilon_S + \epsilon_T + \epsilon_{ST}.$$

The mean value of the interference energy ϵ_{ST} is *nil*, but its mean square is

$$\overline{\epsilon_{ST}^2} = 2\overline{E_S E_S^* \, . \, E_T E_T^*} = 2\bar{\epsilon}_S \epsilon_T.$$

Using this and the relations

$$\overline{\epsilon_S \epsilon_{ST}} = \overline{\epsilon_T \epsilon_{ST}} = 0$$

which are evident, as there is no correlation between the signal and the noise, one obtains

$$\overline{(\epsilon - \bar{\epsilon})^2} = \overline{[(\epsilon_S + \epsilon_T + \epsilon_{ST}) - (\epsilon_S + \bar{\epsilon}_T)]^2} = \overline{\epsilon_T^2} - \bar{\epsilon}_T^2 + 2\epsilon_S \bar{\epsilon}_T.$$

One knows the first two terms from Lorentz's calculation, which gives, in the absence of a signal

$$\overline{\epsilon_T^2} - \bar{\epsilon}_T^2 = \bar{\epsilon}_T^2,$$

hence

$$\overline{(\epsilon - \bar{\epsilon})^2} = 2\epsilon_S \bar{\epsilon}_T + \bar{\epsilon}_T^2,$$

or, as

$$\epsilon_S = \bar{\epsilon} - \bar{\epsilon}_T,$$

$$\overline{(\epsilon - \bar{\epsilon})^2_{\text{cl.}}} = (2\bar{\epsilon} - \bar{\epsilon}_T)\bar{\epsilon}_T. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(I.1)}$$

This is the classical part of the fluctuation, arising from the interference of waves. *The quantum term* is always given by the " law of rare events "

$$\overline{(\epsilon - \bar{\epsilon})^2_{\text{qu.}}} = h\nu\bar{\epsilon} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(I.2)}$$

as if the energy were present in the form of particles (photons), with energy $h\nu$. Adding (I.1) and (I.2), one obtains equation (4).

APPENDIX II.

CLASSICAL ENERGY EXCHANGE BETWEEN A WAVE AND AN ELECTRON BEAM.

In a rectangular wave guide with cross-section $a \times b$ (fig. 5), the components of the electric field are, in the TE_{01} mode,

$$E_x = E_0 \cos\left(\frac{\pi y}{b}\right) \cdot \cos 2\pi\nu\left(t - \frac{z}{V}\right), \quad E_y = E_z = 0,$$

where V is the phase velocity

$$\mathrm{V}=c\left[1-\left(\frac{c}{2b\nu}\right)^{2}\right]^{1\ 2}$$

Thus the mean electric energy is, per unit length of wave guide,

$$\frac{1}{8\pi}\frac{1}{4}\mathrm{E}_0^2\,ab,$$

and the mean total energy, electric and magnetic, is twice as much. This energy moves with the group velocity $\mathrm{U}=c^2/\mathrm{V}$. The mean number $\overline{\mathrm{N}}$ of photons in the information cell is obtained by equating the energy flux expressed by $\overline{\mathrm{N}}$ and expressed by E_0. As the cell occupies a time $\Delta t=1/\Delta\nu$, the flux per second, *i.e.*, the power, is

$$h\nu\overline{\mathrm{N}}\Delta\nu=\frac{1}{8\pi}\,\tfrac{1}{2}\mathrm{E}_0^2ab\mathrm{U}=\frac{1}{16\pi}\,\mathrm{E}_0^2abc\left[1-\left(\frac{c}{2b\nu}\right)^{2}\right]^{1/2}. \quad . \quad . \quad (\mathrm{II}.1)$$

Consider an electron which traverses the guide on the x-axis. Assume that its velocity v is large enough to neglect its change during the transit. Thus the number of quanta exchanged—gained or lost—will be, in the mean,

$$\bar{n}=\frac{e}{h\nu}\overline{\left|\int_{-1/2a}^{1/2a}\mathrm{E}\,dx\right|}=\frac{ev}{h\nu}\overline{\left|\int_{t_0}^{t_0+a/v}\mathrm{E}\,dt\right|},$$

where the mean is to be taken with respect to the entrance time t_0. A simple calculation gives

$$\bar{n}=\frac{2}{\pi}\frac{\sin\theta}{\theta}\frac{ea}{h\nu}\mathrm{E}_0 \qquad (\theta<\pi), \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{II}.2)$$

where θ is half the " transit angle," *i.e.*,

$$\theta=\pi\nu a/v.$$

The condition $\theta<\pi$ for the validity of (11.2) is always satisfied in an efficient arrangement, in fact, one must take $\theta<\frac{1}{2}\pi$.

Eliminating the field amplitude E_0 between (II.1) and (II.2), one obtains a direct relation between the photon number $\overline{\mathrm{N}}$ and the exchange parameter $\bar{n}$:

$$\begin{aligned}\bar{n}^2&=\frac{32}{\pi^2}\frac{\sin^2\theta}{\theta^2}\frac{a}{b}\left[1-\left(\frac{c}{2b\nu}\right)^{2}\right]^{-1/2}\frac{\Delta\nu}{\nu}\overline{\mathrm{N}}\\&=\frac{1}{137}\frac{32}{\pi^3}\frac{\sin^2\theta}{\theta}\frac{v}{\nu b}\left[1-\left(\frac{c}{2b\nu}\right)^{2}\right]^{-1/2}\frac{\Delta\nu}{\nu}\overline{\mathrm{N}}. \quad . \quad . \quad (\mathrm{II}.3)\end{aligned}$$

As explained in the text, the coefficient of $\overline{\mathrm{N}}$ must be made as large as possible. The depth a appears only in the form $\sin^2\theta/\theta$, whose maximum value 0·723 is reached at $\theta=67°$, *i.e.*, at a transit angle of 134°. The width b, on the other hand, appears in the factor

$$1/b\,.\left[1-\left(\frac{c}{2b\nu}\right)^{2}\right]^{1/2}=c/b\mathrm{U}.$$

The smallest admissible value of U is obtained if the group velocity is zero at the lower end of the frequency band, *i.e.*, if frequencies below $\nu=\frac{1}{2}\Delta\nu$ are cut off. In this case

$$b=\tfrac{1}{2}c/(\nu-\tfrac{1}{2}\Delta\nu),$$

hence the best value of the factor in question is the reciprocal of

$$b\left[1-\left(\frac{c}{2b\nu}\right)^2\right]^{1/2}=\frac{c}{2(\nu-\frac{1}{2}\Delta\nu)}\left[\frac{\Delta\nu}{\nu}\left(1-\frac{1}{4}\frac{\Delta\nu}{\nu}\right)\right]^{1/2}\approx\frac{c}{2\nu}\left(\frac{\Delta\nu}{\nu}\right)^{1/2},$$

assuming $\Delta\nu/\nu\ll 1$, *i.e.*, the band to be narrow. Substituting these values, one obtains

$$\overline{n}^2=\frac{1{\cdot}50\,v}{137\,c}\left(\frac{\Delta\nu}{\nu}\right)^{1/2}\mathrm{N}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(II.4)}$$

which is equation (15) of the text.

If repeated passages are used, it is evidently permissible to consider them of equal value, so long as their number P is small. It may be asked when this assumption will lead to an appreciable error. An estimate can be made by introducing the concept of the "instantaneous frequency"* which will vary slowly between the limits $\nu\pm\frac{1}{2}\Delta\nu$. A somewhat long calculation, which may be omitted, gives the result that for large P's the factor of $\overline{n}$ is not exactly P but approximately

$$\left(\overline{\frac{\sin \mathrm{P}\pi\delta\nu/\nu}{\pi\delta\nu/\nu}}\right)=\mathrm{P}\left[1-\frac{1}{6}\mathrm{P}^2\pi^2\left(\frac{\overline{\delta\nu^2}}{\nu}\right)+\ldots\right],$$

where $\delta\nu$ is the deviation of the instantaneous from the mean frequency. Assuming that the instantaneous frequency is uniformly distributed in the available band $\Delta\nu$, one has $\overline{\delta\nu^2}=1/12(\Delta\nu)^2$. In the text, we have assumed that the maximum number of admissible passages is $\mathrm{P}_{\max}=\nu/\Delta\nu$. Even this gives an error of less than 14 per cent, which justifies our approximate treatment.

APPENDIX III.

ENERGY EXCHANGE OF ELECTRONS AND WEAK FIELDS ACCORDING TO QUANTUM MECHANICS.

We apply to the problem the standard perturbation method of wave mechanics. As the beam width can be considered as small in comparison to the spatial periods of the field, it is sufficient to start from Schrödinger's one-dimensional equation with vector potential A_x,

$$\frac{h}{2\pi i}\frac{\partial\Psi}{\partial t}=\frac{h^2}{8\pi^2 m}\frac{\partial^2\Psi}{\partial x^2}-\frac{ieh}{2\pi mc}\mathrm{A}_x\frac{\partial\Psi}{\partial x}. \quad . \quad . \quad . \quad \text{(III.1)}$$

We put

$$\mathrm{A}_x=\mathrm{A}_0\cos\omega_0 t, \qquad \mathrm{A}_y=\mathrm{A}_z=0, \qquad \omega_0=2\pi\nu_0,$$

---

* This concept is due to Helmholtz and has been first applied in communication theory by Balth. v.d. Pol. (1930). *Cf.* also J. R. Carson and T. C. Fry (1937).

which corresponds to an electric field

$$\mathrm{E}_x=\mathrm{E}_0 \sin \omega_0 t, \qquad \mathrm{E}_y=\mathrm{E}_z=0, \qquad \mathrm{E}_0=\mathrm{A}_0/\omega_0 c,$$

ν_0 is here the frequency of the field, which previously we have called ν, but in this calculation we will reserve ν for the frequency of the oscillators by which we will later replace the electron.

Assume that the solutions for zero field are written in the form

$$\Psi_n=\psi_n \exp\left(-i2\pi\epsilon_n t/h\right),$$

where ϵ_n is the energy of the unperturbed electron in some state " n ". The general solution of equation (III.1) can be written in the form

$$\Psi=\sum_m c_m(t)\Psi_m.$$

With this substitution, equation (III.1) becomes

$$\Sigma \frac{dc_m}{dt}\Psi_m=\frac{e}{mc}\Sigma c_m \mathrm{A}_x \frac{\partial \Psi_m}{\partial x}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad \text{(III.2)}$$

Let us assume that the functions $\psi_m(x)$ are orthonormal

$$\int \psi_n^* \psi_m \, dx=\delta_{nm}.$$

Multiplying equation (III.2) by Ψ_r^*, integrating over the whole domain and making use of the orthonomality conditions, we obtain ordinary differential equations for the coefficients $c_r(t)$ in the form

$$\frac{dc_r}{dt}=i\frac{\pi e \mathrm{A}_0}{mch}\sum_m c_m p_{mr}\{\exp\left[i(\omega_0+\omega_{rm})t\right]+\exp\left[-i(\omega_0-\omega_{rm})t\right]\}, \quad . \quad . \quad . \qquad \text{(III.3)}$$

where

$$p_{mr}=\frac{h}{2\pi i}\int \psi_r^* \frac{\partial}{\partial x}\psi_m \, dx$$

is the matrix element of the momentum p of the electron, and

$$\omega_{rm}=\frac{2\pi}{h}(\epsilon_r-\epsilon_m)$$

is the circular frequency which corresponds to the transition $r \rightarrow m$.

Assume now that at the instant t_0 the electron is in the state $m=0$. Thus at the start the coefficient c_0 is unity, all the others are zero. Let us also assume that during the time of interaction, *i.e.*, during the passage of the electron, the other states c_m increase slowly enough to neglect transitions other than $0 \rightarrow r$. Thus we obtain, integrating equation (III.3) for the coefficient c_r at the time t,

$$c_r(t)=\frac{ne\mathrm{A}_0}{mch}p_{0r}\left\{\frac{\exp\left[i(\omega_0+\omega_r)t\right]-\exp\left[i(\omega_0+\omega_r)t_0\right]}{(\omega_0+\omega_r)} - \frac{\exp\left[-i(\omega_0-\omega_r)t\right]-\exp\left[-i(\omega_0-\omega_t)t_0\right]}{(\omega_0-\omega_r)}\right\}.$$

The absolute square of c_r is the probability that after the time $t-t_0$ the electron will be in the state " r ", having absorbed an energy corresponding to the frequency $\omega_{r0}=\omega_r$. Substituting E_0 instead of A_0, we find finally for the probability distribution of the electron at time t the expression

$$c_r c_r^* = \left(\frac{eE_0}{mhv_0}\right)^2 |p_{0r}|^2 \left\{\frac{\sin^2 \frac{1}{2}(\omega_0+\omega_r)(t-t_0)}{(\omega_0+\omega_r)^2} + \frac{\sin^2 \frac{1}{2}(\omega_0-\omega_r)(t-t_0)}{(\omega_0-\omega_r)^2} + 2\cos 2\omega_0 t\,\frac{\sin^2 \frac{1}{2}\omega_r(t-t_0)}{\omega_0{}^2-\omega_r{}^2}\right\}. \quad . \quad \text{(III.4)}$$

There are several interesting features in which this differs from the classical result. Note first, that the expression is even in ω_r, hence an electron has the same probability to gain or to lose a certain amount of energy. Note also that only the last term depends on the phase of the electric field at the instant t_0, the others depend only on the interaction time $t-t_0$. But even this last term has a frequency *twice* that of the electric field. It is rather doubtful whether this corresponds to an experimentally observable effect. Our calculation relates to weak fields, where $\bar{n}$ will be smaller than unity; but in order to observe even $\bar{n}=1$, the electron beam must be monochromatized to such an extent that its entrance phase becomes indefinite within a whole cycle. This is of no importance for our subsequent calculations, as we will consider only the mean value of (III.4), averaged over the entrance times t_0, in which the last term vanishes.

There remains now the problem to apply equation (III.4), which is a well-known result of wave mechanics, to our problem of an electron traversing the oscillating field between two conducting plates. This means calculating the matrix of the (mechanical) momentum p. We will evaluate this by a classical method, making use of the correspondence principle. This is justified in our case, as the de Broglie wavelength of the electrons is always very small with respect to the distance a between the plates.

In order to simplify the calculations, and to avoid singularities, we replace the electron by a plane density wave of total charge e (per unit cross-section)

$$\rho(x, t) = \frac{e}{\sqrt{(\pi)}v\tau}\exp\left[-(t-t_0-x/v)^2/\tau\right].\, D(x). \quad . \quad . \quad \text{(III.5)}$$

$D(x)$ is an " annihilation factor " which is unity inside the guide, *i.e.*, for $-\frac{1}{2}a<x<\frac{1}{2}a$, and zero outside. Apart from this, *i.e.*, if the electron is at more than about $v\tau$ from the walls, the distribution is gaussian. τ is a small time, with which we will go ultimately to the limit zero. It will make the calculations easier if for the start we ignore the factor $D(x)$ and make use of it only towards the end.

We make use of the Fourier formula

$$\exp -[(t-x/v)/\tau]^2 = \sqrt{(\pi)}\tau \int_{-\infty}^{\infty} \exp\left[-(\pi\tau\nu)^2 - 2\pi i\nu(t-x/v)\right] d\nu.$$

This is to say that we decompose the electron into travelling waves of density

$$\rho_\nu \, d\nu = e/v \exp -(\pi\tau\nu)^2 \, . \exp\left[-2\pi i\nu(t-x/v)\right] d\nu.$$

We have put the entrance time $t_0=0$, as it plays no part in the mean values which we want to calculate. In the following formulæ we will drop also the factor $\exp -(\pi\tau\nu)^2$, which cuts off the spectrum at very high frequencies, but which is practically unity in the frequency region to be investigated.

The electric field produced by the electron follows from Poisson's equation

$$\frac{\partial \mathrm{E}}{\partial x} = 4\pi\rho.$$

Let us decompose this too into harmonic components by putting

$$\mathrm{E} = \int_{-\infty}^{\infty} \mathrm{E}_\nu \, d\nu.$$

One obtains

$$\mathrm{E}_\nu = i\frac{2e}{\nu} \exp\left[-2\pi i\nu\left(t-\frac{x}{v}\right)\right]$$

for the progressive field waves which accompany the *free* electron. But in a plane condenser with boundaries at $x=\pm\frac{1}{2}a$, this field induces surface charges

$$\sigma_\nu(\pm\tfrac{1}{2}a) = \mp\frac{1}{4\pi}\mathrm{E}_\nu(\pm\tfrac{1}{2}a) = \mp\frac{ie}{2\pi\nu}\exp\left[-2\pi i\nu\left(t\mp\frac{1}{2}\frac{a}{v}\right)\right],$$

which cut off the field beyond $x=\pm\frac{1}{2}a$.

We can now calculate the electric moment of the charges in the condenser, *i.e.*, of the electron charge and the surface charges, whose sum total is zero. The Fourier component of the moment in the frequency interval $\nu \, . \, d\nu$ is

$$d\nu\left\{\int_{-\frac{1}{2}a}^{\frac{1}{2}a} x\rho_\nu \, dx + \tfrac{1}{2}a[\sigma_\nu(\tfrac{1}{2}a) - \sigma_\nu(-\tfrac{1}{2}a)]\right\} = i\frac{ev}{2\pi^2\nu^2}\sin(\pi\nu a/v) \, . \exp(-2\pi i\nu t) \, d\nu.$$

We now equate this to the "standard oscillator" of electromagnetic theory, *i.e.*, to an electron of charge e which oscillates with a frequency ν around an equal and opposite charge, with an instantaneous amplitude X_ν. This gives

$$\mathrm{X}_\nu d\nu = i\frac{v}{2\pi^2\nu^2}\sin(\pi\nu a/v) \, . \exp(-2\pi i\nu t) \, d\nu.$$

The mechanical momentum of this equivalent oscillator is $m\dot{\mathrm{X}}_\nu \, d\nu$. Thus the momentum is, for the frequency interval ν , $d\nu$,

$$\mathrm{P}_\nu \, d\nu = \frac{mv}{\pi\nu}\sin(\pi\nu a/v) \, . \exp(-2\pi i\nu t) \, d\nu. \quad . \quad . \quad . \quad \text{(III.6)}$$

We cannot introduce this directly into equation (III.4), because that equation relates to discrete, steady states. In that case one has only to take the absolute square of the matrix elements to form the quantities $|p_{0r}|^2$ which figure in equation (III.4). But here we have to do with continuous states, and we must also remember that the oscillators, according to equation (III.6), replace the electron during the finite transit time only. In order to obtain the equivalents of the "oscillator strengths" $|p_{0r}|^2$, we must now multiply the momentum amplitudes $\mathrm{P}_\nu \, d\nu$ with the conjugate amplitudes $\mathrm{P}^*_\mu \, d\mu$ which belong to another interval μ, $d\mu$; we must integrate over the whole domain of the μ, and finally average over the transit time a/v. It is to say that we must calculate, instead of $|p_{0r}|^2$,

$$|p_{0\nu}|^2 \, d\nu = 2 \frac{v}{a} d\nu \int_{-\frac{1}{2}a/v}^{\frac{1}{2}a/v} dt \int_{-\infty}^{\infty} \mathrm{P}_\nu \mathrm{P}^*_\mu \, d\mu. \quad . \quad . \quad . \quad \text{(III.7)}$$

We had to add the factor two, because in wave mechanics ν is considered as a positive quantity, while in the classical theory it runs from $-\infty$ to $+\infty$. Substituting from (III.6),

$$|p_{0\nu}|^2 = 2 \left(\frac{mv}{\pi}\right)^2 \frac{v}{a} \frac{\sin \pi\nu a/v}{\nu} \int_{-\frac{1}{2}a/v}^{\frac{1}{2}a/v} dt \int_{-\infty}^{\infty} \frac{\sin \pi\mu a/v}{\mu} \exp\left[2\pi i(\mu-\nu)t\right] d\mu.$$

After integration with respect to t, the double integral becomes

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin \pi\mu a/v}{\mu} \sin \frac{\sin \pi(\mu-\nu)a/v}{\mu-\nu} d\mu.$$

With the substitutions $x = (2\mu/v) - 1$ and $\theta = \pi\nu a/v$, this is transformed into

$$\frac{2}{\pi\nu} \int_{-\infty}^{\infty} \frac{\sin \frac{1}{2}\theta(1+x) \, . \, \sin \frac{1}{2}\theta(1-x)}{1-x^2} dx = \frac{2}{\pi\nu} \int_0^{\infty} \frac{\cos \theta x - \cos \theta}{1-x^2} dx = \frac{1}{\nu} \sin \theta,$$

thus finally

$$|p_{0\nu}|^2 \, d\nu = 2 \left(\frac{mv}{\pi}\right)^2 \frac{v}{a} \frac{\sin^2 \theta}{\nu^2} d\nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(III.8)}$$

This means simply that during the transit time a/v the coherence region of any frequency ν is $\nu \pm v/2a$, with a bandwidth v/a, which is the reciprocal of the transit time. A frequency ν cooperates with any frequency μ inside this band in producing the oscillator strength.

Substituting (III.8) instead of $|p_{0r}|^2$ into equation (III.4), where we drop the last term whose mean value is zero, and replacing $t-t_0$ by the transit time a/v, we obtain

$$|c_\nu|^2 \, d\nu = \frac{1}{h} \left(\frac{2\pi e^2}{hc}\right) \frac{\mathrm{E}_0^2 acv}{4\pi^3 \nu_0^2} \frac{d\nu}{\nu^2} \sin^2 (\pi\nu a/v) \left\{ \frac{\sin^2 \pi(\nu_0+\nu)a/v}{[(\nu_0+\nu)a/v]^2} + \frac{\sin^2 \pi(\nu_0-\nu)a/v}{[(\nu_0-\nu)a/v]^2} \right\} \quad . \quad . \quad . \quad \text{(III.9)}$$

This is the probability for an electron to be, after the passage, in a state ν, $d\nu$, that is to say, to have absorbed the energy $h\nu$. Thus the energy spectrum of the originally monokinetic electrons after the passage through the field is of the form

$$\mathrm{S}(\theta) = \frac{\sin^2 \theta}{\theta^2} \left\{ \frac{\sin^2 (\theta+\theta_0)}{(\theta+\theta_0)^2} + \frac{\sin^2 (\theta-\theta_0)}{(\theta-\theta_0)^2} \right\}, \quad . \quad . \quad \text{(III.10)}$$

where the frequencies (and energies) are expressed by the half transit-angles

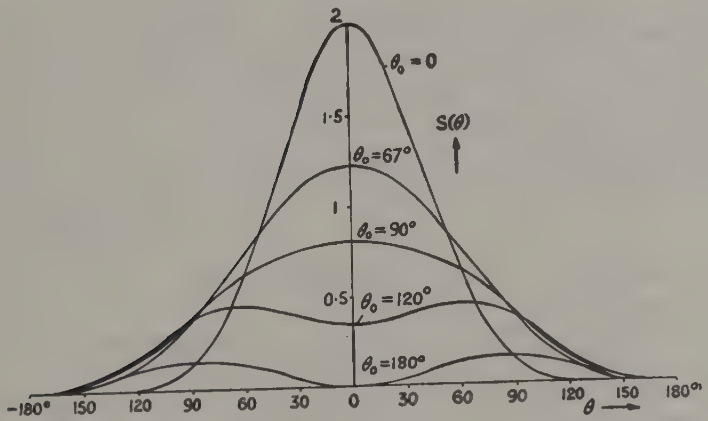
$$\theta = \pi\nu a/v, \qquad \theta_0 = \pi\nu_0 a/v.$$

This function is shown in fig. 8 for a few values of θ_0. Experimental *checking*, though difficult, may not be impossible (*cf.* Macdonald and Kompfner 1949).

The mean energy exchanged with the field, lost or gained, is by (III.10)

$$\bar{n}h\nu_0 = \int_0^\infty h\nu\,|c_\nu|^2\,d\nu = \frac{1}{137}\,\frac{E_0^2 acv}{4\pi^3\nu_0^2}\,F(\theta_0), \quad . \quad . \quad \text{(III.11)}$$

Fig. 8.



The function $S(\theta)$, equation (III.10), which gives the energy distribution of electrons after the transit. Gains and losses have equal probabilities. The "resonance point" at which the electron has gained or lost one quantum $h\nu_0$ of the oscillating field is at $\theta=\theta_0$.

where $F(\theta_0)$ is the integral

$$F(\theta_0) = \int_0^\infty \frac{d\theta}{\theta}\left[\frac{\sin^2(\theta+\theta_0)}{(\theta+\theta_0)^2} + \frac{\sin^2(\theta-\theta_0)}{(\theta-\theta_0)^2}\right].$$

For $\theta=67°$, which was the most favourable case in the classical theory, the value of the integral is found to be 1·22. Substituting this into equation (III.11) and expressing, as before the field E_0 by the number of photons, $\bar{N}$, by means of equation (II.1), and finally giving the wave guide the same width b which was found the best in the classical case, one finds

$$\bar{n} = \frac{1 \cdot 0}{137}\,\frac{v}{c}\left(\frac{\Delta\nu}{\nu}\right)^{1/2}\bar{N},$$

where we have again written ν for the field frequency, instead of ν_0. This is the energy exchange in weak fields, discussed in the text.

APPENDIX IV.

CONDITIONS FOR OPTIMUM EXPLORATION OF ELECTRIC FIELDS BY ELECTRONS.

As explained in the text, we adopt as the criterion of optimum measurement the condition that

$$\frac{\overline{(nM-CN^{1/2})^2}}{(\overline{nM})^2},$$

must be as small as possible. The proportionality factor C is given by the condition

$$\overline{(nM-CN^{1/2})}=0,$$

or, as we are dealing here only with the "classical" case, in which the fluctuations are relatively small, $C=\overline{nM}\,.\,(\overline{N})^{-1/2}$. As the fluctuations of n, the energy exchange parameter, and of M, the electron number, can be supposed as independent of one another, we can write $\overline{nM}=\bar{n}\,.\,\overline{M}$.

Let us now first calculate

$$\overline{n^2M^2}=\overline{[(\bar{n}+\delta n)(\overline{M}+\delta M)]^2}=\overline{(\bar{n}\overline{M}+\bar{n}\delta M+\overline{M}\delta n)^2}=\bar{n}^2\overline{M}^2+\bar{n}^2\overline{\delta M^2}+\overline{M}^2\overline{\delta n^2}$$

$$=\bar{n}^2\overline{M}^2+\bar{n}^2\overline{M}+\overline{M}^2\bar{n}.$$

Here we have neglected the flunctuation term of the highest order, and we have assumed

$$\overline{\delta n\delta M}=0, \qquad \overline{\delta n^2}=\bar{n}, \qquad \overline{\delta M^2}=\overline{M}.$$

The first is the assumption of the relative independence of the fluctuations of n and M. The second is the assumption that the absorption of a quantum is independent of those previously absorbed or lost, which is certainly admissible so long as the electron velocity is not changed appreciably. The third assumption means shot effect without space charge smoothing. This is a logical assumption in a theory in which each electron is considered to interact singly with the field. With these formulæ, we can now calculate

$$\overline{(nM-CN^{1/2})^2}=\overline{n^2M^2}-2C\overline{(\bar{n}+\delta n)(\overline{M}+\delta M)(\overline{N}^{1/2}+\tfrac{1}{2}\delta N(\overline{N})^{-1/2})}$$

$$+\overline{C^2(\overline{N}^{1/2}+\tfrac{1}{2}\delta N(\overline{N})^{-1/2})^2}=(\bar{n}\,.\,\overline{M}-C\overline{N}^{1/2})^2+\bar{n}^2\overline{M}+\overline{M}^2\bar{n}+\tfrac{1}{4}C^2\,\frac{\overline{\delta N^2}}{\overline{N}}, \quad . \;\; . \;\; . \quad \text{(IV.1)}$$

where the first term is zero.

In order to calculate the last term, we consider the fluctuation δN of the photons as the sum of two independent components δN_1 and δN_2

$$\delta N=\delta N_1+\delta N_2,$$

of which the first represents the "natural" fluctuations, subject to the law $\overline{\delta N_1^2}=\overline{N}$, and the second is due to the exploring electrons, *i.e.*, it is the excess of emitted over absorbed photons. Let M_1 be the number of beam electrons which pass through the field in a retarding phase, M_2 the number of accelerated electrons. As in the mean $\overline{M}_1=\overline{M}_2=\frac{1}{2}\overline{M}$, and as in the mean each electron produces or absorbs $\bar{n}$ photons, we can write

$$\delta_2 N=\bar{n}(M_1-M_2)=\bar{n}(\delta M_1-\delta M_2),$$

and

$$\overline{\delta_2 N^2}=\bar{n}^2(\overline{\delta M_1^2}+\overline{\delta M_2^2})=\bar{n}^2(\overline{M}_1+\overline{M}_2)=\bar{n}^2\overline{M},$$

where we have assumed $\overline{\delta M_1 \delta M_2}=0$, in accordance with our assumption of a normal, entirely random shot effect.

We now substitute these results into equation (IV.1), and replacing C^2 by its value $\bar{n}^2\overline{M}^2/\overline{N}$, we obtain for the mean square relative error in the measurement of the amplitudes

$$\frac{\overline{(nM-CN^{1/2})^2}}{\overline{nM}^2}=\frac{1}{4\overline{\overline{N}}}+\frac{1}{\bar{\bar{n}}}+\bar{n}\left(\frac{1}{\overline{\overline{nM}}}+\frac{\overline{nM}}{4\overline{\overline{N}}^2}\right). \qquad \text{(IV.2)}$$

This is equation (24), discussed in the text.

It is not so easy to give an exact analysis of the errors in the measurement of phase, because devices for the running measurement of phase, with uniform error, are by no means simple. But the order of magnitude can be seen immediately if we apply the uncertainty principle to the exploring electrons, in the form

$$\delta\epsilon\delta t \approx 1.$$

δt is the uncertainty in the time at which the electron passes some fixed plane, *e.g.*, a fluorescent screen on which the waveform is recorded. The uncertainty in the energy ϵ is, assuming that the electrons were monokinetic at the start,

$$\delta\epsilon \doteq h\nu\delta n,$$

n being the number of quanta which it may have absorbed from or lost to the field. But as δn is of the order $\bar{n}^{1/2}$, we obtain

$$\nu\delta t \approx (\bar{n})^{-1/2}.$$

In words, the oscillogram of the signal is traced with a spot whose mean width in the direction of time is a fraction $1/\sqrt{\bar{n}}$ of a cycle. Thus one can determine the phase with the same order of relative error as the amplitude.

CIV. *Density Variations in Aluminium.*

By J. L. SNOEK,
Philips Research Laboratories,
Eindhoven*.

[Received August 25, 1950.]

SUMMARY.

From careful determinations of the density of samples of aluminium in various conditions the conclusion is drawn that grain boundaries in aluminium may have a more spacious structure than is generally assumed.

COLD WORK AND DENSITY.

A priori considerations lead one to assume that the density of any metal, particularly however, that of metals with a close packed lattice structure will diminish as a result of cold work.

Various authors have found such a decrease (Burgers), but very careful work by Smart, Smith and Phillips on copper has shown that the decrease observed becomes less and less as the amount of oxygen in the lattice diminishes so that in all probability in extremely pure copper there is only a very small effect of cold work on the density ($<0{\cdot}01$ per cent).

The above serves to show that one has to be careful in drawing conclusions from results obtained on so-called pure metals. The only satisfactory procedure seems to be to vary the amount of a certain impurity and extrapolate the observed data towards concentration zero. In the majority of cases this has not been done and the results obtained remain open to serious doubt.

DENSITY AND NUMBER OF VACANCIES.

Our measurements on the density of aluminium in various conditions were started with the object of obtaining some direct data on the concentration of vacancies in this metal, the underlying idea being that in thermodynamical equilibrium this concentration would depend very much on the temperature and that it would be possible to " freeze in " a certain number of vacancies by cooling rapidly from a high temperature, or, if the process of annihilation were very slow to eliminate them by a process of cold work and annealing.

Aluminium as an object for research was preferred to gold and silver or copper on account of its light weight, which makes the measurement of the density roughly four times more accurate than is possible for the heavier elements. Magnesium, which is even more favourable in that respect, was discarded because of its non-cubic structure and the difficulty in cold working it satisfactorily.

---

* Communicated by the Author.

Before describing the results, which were very unexpected, we shall give some details on the technique of the measurement and on the chemical composition of the samples.

TECHNIQUE OF THE MEASUREMENT.

Determination of the density on aluminium foil of half a millimeter thickness showed us—as was mentioned to us by Dr. W. Ch. van Geel of this laboratory before—that the surface absorbs variable amounts of water vapour. It is desirable, therefore, to make the surface play an unimportant part in the determination.

The analytical balance used—a new one, capable of an accuracy of 0·1 milligram—allowed of a maximum weight of 200 gram. As the standard shape of the object of which the density had to be determined we choose accordingly a cylinder with a diameter of 19 mm. and a length of 85 mm., having at one end a butt carrying a hole through which a molybdenum wire of 50 micron thickness was passed, serving to suspend the sample on one side of the balance.

The determination of the "dry" weight of the sample to within 0·1 milligram presented no undue difficulties, the weights used having been calibrated and the balance being placed in a room kept at constant temperature.

The determination of the weight in water presents more difficulties. Doubly distilled water was placed in a massive block made of aluminium containing four channels, two vertical ones stopping above the bottom of the block and two horizontal ones connecting the vertical holes. The vertical holes had diameters of 50 mm. subs 20 mm. each.

The wider vertical hole contained the sample, the smaller hole a Beckmann thermometer allowing the temperature of the water to be determined to within a few thousandths of a degree centigrade. The temperature never deviated more than 2 degrees from the temperature to which all density determinations were reduced, viz. 20° C. The usual corrections were of course applied.

The sample as a rule was turned on the lathe only a few hours before a determination was made, but measurements taken some days later showed no systematic deviations from this first value. Sharp corners were always avoided and the surface turned to a maximum of smoothness. The main difficulty lies in the capillary depression around the molybdenum wire. *Oleum pedis bovis* was finally chosen as a suitable means of reducing the uncertainties in the capillary behaviour to a minimum.

The question whether a given determination is satisfactory from this point of view is answered by a determination of the apparent sensitivity of the balance under the conditions of the experiment. The adding or taking off of one milligram, should always cause the same displacement of the needle. If it does not, something is wrong with the wetting of the molybdenum wire.

Heat-treatments as a rule were carried out on pieces of aluminium bar diameter 20 mm. as taken from stock. After the treatment the sample was given a suitable shape for a determination of the density. Such determinations are called a-measurements. For comparison an additional heat-treatment was sometimes applied to a cylinder which had already served for a measurement of the density. On account of the oxide-layer formed it was then first turned off on the lathe a second time. Measurements made on such a smaller specimen are called b-measurements.

IMPURITIES IN THE ALUMINIUM USED.

Aluminium was obtained in the form of long bars having a diameter of 20 mm. and a Vickers hardness of 29·2. Chemical analysis revealed the presence of the following impurity contents :

Impurities in aluminium used

Element	Per cent by weight
Si	0·006
Cu	0·003
Ga	0·002
Fe	0·004

The main impurity, however, is Mg, the amount of which was estimated from spectrochemical data to be at least 0·025 per cent by weight. No decrease in the Mg-content was observed after heat-treating the aluminium for many hours at 600° C.

TABLE I.

Number of specimen	Heat-treatment (in air)	$V_{\alpha ph}$	ρ_{20}		Mean grain-size (mm.)
1 (a)	As received	29·2	2·69885	85	—
2 (a)	,, ,,		85		
3 (a)	50 h. 400° C.	13·7	57	58	—
4 (a)	,, ,,		59		
5 (a)	,, ,, +2 h. 600° C.	14·4	81	84	—
6 (a)	,, ,, ,,		86		
7 (a)	,, ,, +20 h. 600° C.	13·3	78	86	—
8 (a)	,, ,, ,,		95		
9 (a)	,, ,, +200 h. 600° C.	13·6	89	92	—
10 (a)	,, ,, ,,		96		
1 (b)	50 h. 400° C.		40	42	0·5
2 (b)	,, ,,		43		
11 (a)	,, ,,		60	55	0·5
12 (a)	,, ,,		50		
9 (b)	quenched from 600° C. in water		90	95	6
10 (b)	quenched from 600° C. in water		100		
3 (b)	50 h. 400° C.+200 h. 600° C.		78	78	4
4 (b)	,, ,, ,, ,,		78		
5 (b)	,, ,, ,, ,,		92	94	4
6 (b)	,, ,, ,, ,,		95		
7 (b)	,, ,, ,, ,,		101	96	5
8 (b)	,, ,, ,, ,,		90		

EXPERIMENTAL RESULTS.

The outcome of our experiments was at first confusing and at any rate not in accordance with what we expected to find. It was found that heating at increasing temperatures caused the density to increase (Table I.) whereas quenching on the whole had no effect upon the measured density. The idea occurred to us that this might be an effect of the grain size on the mean density.

A second series of experiments was then started in which the temperature of the heating was kept constant at 600° C., the times of heating being varied between wide limits This led to essentially the same result (Table II.).

TABLE II.

Number of specimen	Heat-treatment	ρ_{20}	Mean grain-size (mm.)
13 (a)	2 min. 600° C. (salt bath)	2·69841	<1
14 (a)	5	76	1
15 (a)	10	68	2
16 (a)	25	76	2
17 (a)	50	80	3
18 (a)	100	81	3

TABLE III *.

(Fine grained as cast.)

Specimen	Heat-treatment	ρ_{20}		Mean grain-size (mm.)
A	As received	2·69975	$\Delta=+0{\cdot}00189$	<1
A	2 h. 600° C.	2·70164		6
B	As received	2·69981	$\Delta=+0{\cdot}00045$	<1
B	2 h. 600° C.	2·70026		4

As a third approach we had a cast made from a fairly cold melt, chilling it rapidly so that a fine grained structure was obtained. This material also became very coarse after a prolonged heating at 600° C. and by this treatment again became denser (Table III.). The aluminium used for this experiment came from another stock and contained less magnesium (0·005 per cent).

DISCUSSION OF RESULTS.

The above-mentioned results are hard to explain. One may first of all think of a spurious effect caused by impurities present in the material. The pursuit of this thought leads, however, to no tangible results, in view of

---

* The circumstance that the density of the cast samples on the whole is somewhat larger than for the extended bars may be due to the fact that it contained less magnesium.

the three very different series of heat-treatments applied to the aluminium, which all lead to essentially the same result. The cast sample besides contained considerably less magnesium than the wrought samples.

A second possibility is the appearance and gradual disappearance again of mysterious holes or gaps. We did not find a trace of such gaps on micro-etched samples.

Somewhat reluctantly we arrive therefore at the conclusion that grain boundaries in aluminium have a much more spacious structure than is generally supposed. They might, for instance, act as "retainers" for a number of vacancies which are formed at high temperature. No specific picture can be offered at the present time.

It is hoped that our experiments will be repeated by other physicists with a view to obtaining an independent check of the result presented here.

ADDENDUM.

These results have been discussed with Professor N. F. Mott and Dr. F. C. Frank who point out that if the density changes are attributed to an open structure at the grain boundary they imply that each boundary is equivalent to a void layer $\frac{1}{3}(5\times10^{-2}\times2\times10^{-4})$ cm$=330$ Å thick, which is hard to believe. It is almost as difficult to attribute the changes to precipitation of solutes, for if the whole of the analytically ascertained impurity were going into and coming out of solution, it would have to be changing volume by about 50 per cent. They have suggested that the aluminium specimens possibly contained gas (most probably hydrogen) which goes into solution at 600° C., stays in under ordinary cooling and particularly with water quench, but comes out again within 50 hours at 400° C. in the form of gas bubbles which lower the density.

This explanation cannot be entirely excluded, especially with regard to those samples (the majority) which came from a stock not prepared in the laboratory. The third series of measurements were carried out on a cast which had been made in the laboratory under the exclusion of free hydrogen, and in the presence of only very little water vapour. It seems unlikely that this cast would contain hydrogen in appreciable quantity. No holes were seen in microscopic examination of specimens after electropolishing, though one would expect to see them if gas bubbles were present to the extent required (one bubble of 30μ, 100 of 3μ or 10^6 of $0{\cdot}3\mu$ diamenter per grain).

Whatever the cause, the inherent difficulties connected with experiments of this kind on aluminium, and the desirability of an independent check, are emphasized.

REFERENCES.

BURGERS, W. G., *Handbuch der Metallphysik*, III **2,** 101.

SMART, J. S., SMITH, A. A., and PHILLIPS, A. J., 1941, *Amer. Inst. Min. Met. Eng.*, **143,** 272.

CV. CORRESPONDENCE.

Radiative Corrections in the Non-Relativistic Limit.

By W. Thirring, D.Phil.,
Dublin Institute for Advanced Studies *.

[Received September 10, 1950.]

In calculating radiative corrections to the Klein–Nishina formula it has been assumed by Corinaldesi and Jost (1948) and by Heitler and Ma (1949) that the corrections vanish in the non-relativistic limit. All terms proportional to the second-order matrix element were interpreted as renormalization, which means defining the electron-charge by the Thomson formula. The question is now, whether Dyson's (1949) renormalization procedure, the only one of which the consistency has been proved, leads to the same result, or whether there are finite corrections proportional to the Thomson cross-section analogous to the case of the magnetic moment. It has been shown by Schafroth (1949), by direct calculation, that after renormalization the fourth-order matrix element vanishes for zero momentum of the incident photon. We will give a simple proof that this holds in all approximations.

One obtains the graph of the nth order to the Compton effect by inserting successively the two external photon lines in any of the electron lines of the electron self-energy graphs of order $n-2$. An electron self-energy graph consists of the main electron line and some closed electron lines and internal photon lines. If one inserts the two external photon lines into the main electron line (energy-momentum $p\mu$) for the limit of zero momentum of the photon, using

$$\frac{\partial}{\partial p_\mu}(\mathbf{p}-m)^{-1}=-(\mathbf{p}-m)^{-1}\gamma_\mu(\mathbf{p}-m)^{-1}\,; \quad \mathbf{p}=p_\mu\gamma^\mu \quad . \quad . \quad (1)$$

and summing over all insertions and all self-energy graphs, one obtains

$$(\mathbf{p}-m)\left[e^1_\mu e^2_\nu \frac{\partial}{\partial p_\mu}\frac{\partial}{\partial p_\nu}\mathrm{S}'(p)\right](\mathbf{p}-m) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

for the matrix element. After renormalization S$'(p)$, as defined by Dyson, has the form

$$\mathrm{S}'(p)=(\mathbf{p}-m)^{-1}[1+\mathrm{B}(p^2-m^2)]+m^{-1}\mathrm{C}(p^2-m^2)\,\dagger, \quad . \quad . \quad (3)$$

---

[*] Communicated by the Author.

[†] Dyson's renormalization casts contributions from proper self-energy parts into the form (3) whilst on multiplying the proper self-energy parts the contribution from the general self-energy part will be

$$(\mathbf{p}-m)^{-1}[1-\mathrm{B}-\mathrm{C}m^{-1}(\mathbf{p}-m)]^{-1}$$

but one readily sees that this is again of the form (3).

where $B(0)=0$. Inserting into (2) we get

$$(2)=[\mathbf{e}^1(\mathbf{p}-m)^{-1}\mathbf{e}^2+\mathbf{e}^2(\mathbf{p}-m)^{-1}](1+B(0))-2[\mathbf{e}^1(pe^2)+e^2(pe^1)]B'(0)$$
$$+4(e^1p)(e^2p)[(\mathbf{p}-m)B''(0)+(\mathbf{p}-m)^2m^{-1}C''(0)].$$

Everything but the uncorrected matrix element vanishes since one can choose for the potentials of the external photons a gauge system in which $(pe^1)=(pe^2)=0$.

Contributions from the closed loops will have the form

$$\int dq\, tr \prod_i (\mathbf{q}+\mathbf{k}_1-m)^{-1}e_i,$$

where k_i are some energy-momentum fourvectors, over some of which one might have to integrate. Inserting one external photon line successively in all electron lines one gets for zero momentum

$$-e_\mu\int dq\,\frac{\partial}{\partial q_\mu}\, tr \prod_i (\mathbf{q}+\mathbf{k}_i-m)^{-1}e_i,$$

which vanishes on partial integration, provided that the integral converges. Otherwise one has to apply regularization in order to avoid contributions from the boundary.

It is evident that the same argument applies to a case where the electron interacts any number of times with the electromagnetic field, and that therefore in all those cases the radiative corrections vanish in the non-relativistic limit.

References.

Corinaldesi, E., and Jost, R., 1948, *Helv. Phys. Acta*, **21**, 183.
Dyson, F. J., 1949, *Phys. Rev.*, **75**, 1736.
Heitler, W., and Ma, S. T., 1949, *Phil. Mag.*, **40**, 651.
Schafroth, M. R., 1949, *Helv. Phys. Acta*, **22**, 501.

Multiple Processes in the Production of Pairs of Electrons by γ-Radiation.

By J. E. Hooper and D. T. King,
H. H. Wills Physical Laboratory*.

[Received September 28, 1950.]

[Plate XXXV.]

It is shown by Heitler (1944) that the cross-section for the simultaneous emission of n quanta by a multiple Bremsstrahlung process is, at high energies, of the order $Z^2r_0^2/137^n$. On this basis one would expect the cross-section for the analogous case of the production of n pairs of electrons by a γ-ray in the electrostatic field of a nucleus to be of the same order of magnitude.

---

* Communicated by Professor C. F. Powell.

We have recently made an investigation of the pairs of electrons created in photographic emulsions exposed at an altitude of 68,000 ft., and resulting from the passage of γ-radiation arising in the decay of π^0-particles. Among 1380 pairs we have found two examples which appear to correspond to the simultaneous production of two pairs of electrons. A mosaic of photo-micrographs of one of these events is shown in Pl. XXXV. In both cases the associated tracks are of sufficient length to allow us to determine the grain-density in the tracks and the momenta of the particles by the scattering method. All the particles have a specific ionization equal to the minimum value for a particle of charge $|e|$, and the total energies of the particles in the two events are 220 and 2000 MeV. respectively.

The energies of the individual particles are shown in the Table.

	Event (A)	Event (B)
E_{1-4} (MeV.)	40±5, 95±10, 60±8, 22± 4	8±3, 960±150, 20±4, 980±150
E_T (MeV.)	220±25	2000±300

These measurements show that nearly all the particles are less massive than protons. Further, in one case two and in the other, three of the particles are less massive than μ-mesons. These facts, and the observed even number of secondary particles, allow us to exclude the possibility that either of the events is due to the collision of a fast neutron with a proton in the emulsion in which a shower of π-mesons is created. We therefore attribute them to the production of pairs of electrons and positrons.

The possibility remains that the events are due to a process first observed by Bradt and Peters (1950) and by Occhialini (1949), in which one of the electrons of a pair, created in the emulsion by a γ-ray, subsequently generates a second pair in the field of a nucleus. We have observed 52 examples of pair creation by a fast particle—the so-called " tridents "—and it would be possible to explain the present observations in terms of a similar occurrence in which the production of the second pair has occurred, by chance, very close to the point of origin of the " primary " pair. Such an event would be difficult to distinguish from one involving the simultaneous production of four electrons in the field of a nucleus, especially in the case of pairs of small angular divergence such as are commonly produced by γ-rays of great quantum energy. We believe, however, that the following argument allows us to exclude this possibility :—

The number of observed events in which a " trident " is found to be associated with a pair of electrons enables us to estimate the average " mean free path " in the emulsion for trident production. The mean value is about 100 cm. The probability of the trident occurring within 10 μ of the point of origin of the pair is therefore 1 in 10^4; and the probability of observing two such events among 1300 pairs is about 1 in 10^5. These values are so small that we can disregard this explanation of the origin of the two events under discussion. Further, the probability that, by chance, two unrelated pairs should be superposed is less than 1 in 10^{12} per cm.^{2} of the emulsion in the conditions of the experiment. We therefore regard our observations as providing strong evidence for the multiple production of pairs of electrons by a γ-ray quantum in the field of a nucleus, the probability of the process being of the order of magnitude suggested by quantum-electrodynamics.

The observations on the " tridents " associated with pairs of electrons will enable us to determine the variation with energy of the cross-section for the production of " pairs " by fast electrons. Preliminary results suggest that there is a rapid increase in the value of the cross-section for the process as the energy of the incident electron increases. We may direct attention to the fact that some caution should be observed in observations of this type because of the possibility of confusion due to "Bremstrahlung". A fast electron loses energy in inelastic collisions which result in the creation of quanta. Especially in the case of electrons of high energy, these quanta may be emitted in directions nearly coincident with the line of motion of the recoiling electron. If, further, such a quantum produces a " pair ", the resulting electrons may appear to arise at a point nearly coincident with the track of the original electron. Such events can be distinguished only with difficulty from true trident production. The possibility of this source of confusion was first pointed out to us by Professor Peierls at the Edinburgh Conference in 1949.

REFERENCES.

HEITLER, 1944, *Quantum Theory of Radiation*, 2nd Ed., p. 181 (Oxford University Press).

BRADT and PETERS, 1950, *H.P.A.*, **24**, 23.

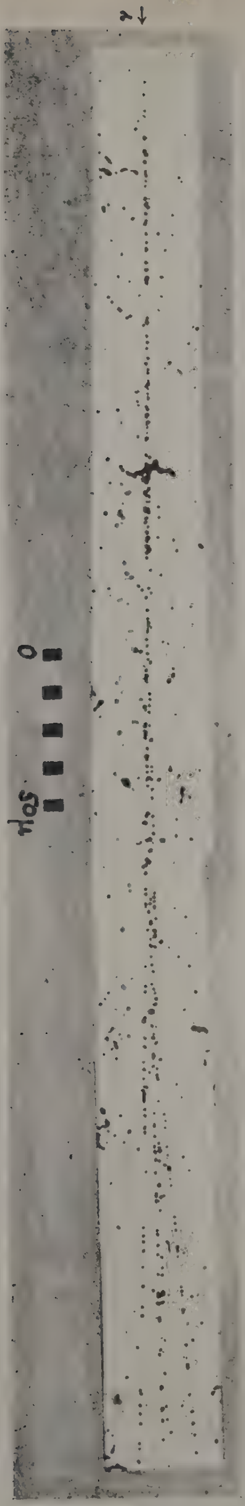
OCCHIALINI, 1949, *Como Conference Reports*.

CVI. *Notices of New Books and Periodicals received.*

Practical Applications of Spectrum Analysis. By H. DINGLE. [Pp. 245.] 14 plates. (Chapman and Hall, 1950.) Price 40*s*. net.

IN a general conclusion to this useful book the author states that in his experience more than half of the spectrographic analyses which have been published during the last thirty years are untrustworthy. Used in conjunction with a standard set of wave-length tables, this book should do much to remedy this state of affairs.

No attempt is made to describe the theoretical laws of spectroscopy, but the relevant aspects of the mechanism of radiation, of sources of luminosity and of the optical principles of the spectrograph are dealt with in so constructive a fashion as to lead to a confident and critical choice and use of the most appropriate apparatus for a very wide range of spectrographic analyses.



A mosaic of photo-micrographs illustrating an event in which two pairs of electrons appear to be created simultaneously by a 220 MeV. γ-ray.

The chapters on the measurement and interpretation of the spectrograms obtained in these analyses, together with the comprehensive tables of Raies Ultimes and the wealth of critical comment on special points of interpretation, taken element by element, embody great experience and form a valuable guide to the less experienced and an indispensable reference work to the more experienced workers in this field.

The writing throughout is clear in style, and the general level of production excellent, except for the reproduction of the plates, the quality of which is somewhat disappointing for a book at this price. D. McG.

Progress in Coal Science. Edited by D. H. Bangham. [Pp. 456.] (Butterworths Scientific Publications, London, 1950.)

This book comprises twenty-six review articles, many of which have appeared in the Monthly Bulletin of the British Coal Utilisation Research Association during the years 1943–1947. They are divided into the following sections :—Modern Experimental Techniques, Fine Particles, Constituents of Coal, Organic Chemistry of Coal Products, and Chemical Aspects of Combustion and Gasification.

The first section contains articles on crystallography, absorption spectrography, mass spectrometry, and chromatography, all as applied to chemical analysis. These should be of use to chemists generally, inside and outside the Coal industry, as also should the articles on the production and measurement of fine particles. The third section contains among others an account by Prof. V. M. Goldschmidt of his work on the rare elements in coal ash, and a review by Dr. C. A. Seyler of his work on coal petrology. In the fourth section is a translation of a German article by O. Gruber on " The Progress in Coal Tar Research in the period 1910 to 1940," and a review of the Fischer-Tropsch reaction. The last section contains an article by Prof. D. T. A. Townsend, which gives a general account of producer gas research, and more specialised articles on fuel beds, combustion kinetics, and other topics.

The book should be most valuable to all scientists in the coal and allied industries and, as indicated, some of the articles should find a wide appeal.

D. D. E.

Astronomy of Stellar Energy and Decay. By Martin Johnson. [Pp. 216, Plates 4.] (Faber and Faber, 1950.) Price 18*s*. 0*d*.

This book, which aims at bridging the gap between popular astronomy and the research literature of stellar physics, can be strongly recommended. It is deliberately set in two parts. The first, though framed for the general intelligent reader, can also be read with interest by students of wider scientific training. Part II. is labelled " student's introduction to analysis of the facts and argument of the theories." As its title suggests it is more qualitative, though never heavily mathematical, in its treatment. It supplements Part I. and serves as a very useful introduction to more advanced treatises. The two parts together can be specially recommended for private reading at the university degree stage. A. M. T.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

CVII. *The Radioactivity of* F^{20}.—I.

By J. V. Jelley*,
Cavendish Laboratory, Cambridge†.

[Received August 23, 1950.]

Summary.

The beta and gamma radiation emitted by F^{20} has been studied by absorption, coincidence and spectrometer techniques. The upper limit of the beta particle spectrum was found to be $5{\cdot}03 \pm 0{\cdot}05$ MeV. and gamma rays of $1{\cdot}63 \pm 0{\cdot}02$ MeV. and $2{\cdot}45 \pm 0{\cdot}06$ MeV. with an intensity ratio of $(8{\cdot}4 \pm 2{\cdot}6)$ to 1 were also observed. No gamma–gamma coincidences were found, and it is concluded that the predominant mode of disintegration of F^{20} leads to a 1·63 MeV. excited state of Ne^{20}. The total energy release $F^{20}-Ne^{20}$ is $6{\cdot}66 \pm 0{\cdot}05$ MeV., and the mass of F^{20} derived from this value is compared with that obtained from the F^{19} (dp) and F^{19} $(n\gamma)$ reactions.

§ 1. Introduction.

The radioactive nucleus F^{20}, of 12 seconds half-life, decays into Ne^{20} with the emission of beta and gamma radiations. Measurements of the energies of these radiations and observations of beta–gamma and gamma–gamma coincidences suggested the decay scheme shown in fig. 1. (Hornyak and Lauritsen 1948, Mattauch and Flammersfeld 1949.)

According to this scheme the emission of a beta-spectrum with an upper limit of 5·1 MeV. leads to the formation of Ne^{20} in an excited state of 2·2 MeV. from which gamma-ray transitions take place either directly to the ground state or via an intermediate state of about 1·6 MeV. Though the lower energy gamma-rays were known to be present (Bower and Burcham 1939, Curran and Strothers 1940), their energy was only inferred from evidence for an excited state of Ne^{20} at 1·6 MeV. obtained in observations on the neutron groups from the reaction F (dn) Ne^{20} (Bonner 1939, and Powell 1940). The total energy release in the decay process on the assumption of this decay scheme is in fair agreement with that derived from the energy release in the reaction F^{19} (dp) F^{20} (Bower and Burcham 1939).

The present experiments were undertaken to improve the energy measurements, to obtain more precise information on the gamma–gamma coincidences, to determine the branching ratio for the two alternative de-excitation modes of the 2·2 MeV. level of Ne^{20} and to detect the gamma-ray of energy ~0·6 MeV. which must be present according to this decay scheme. In the course of the work it appeared that the decay scheme shown in fig. 1 (decay scheme I.) was incorrect and a revised scheme is presented, together with new measurements of the beta- and gamma-ray energies. A new determination of the energy release in the reaction F^{19} (*dp*) F^{20} was also made.

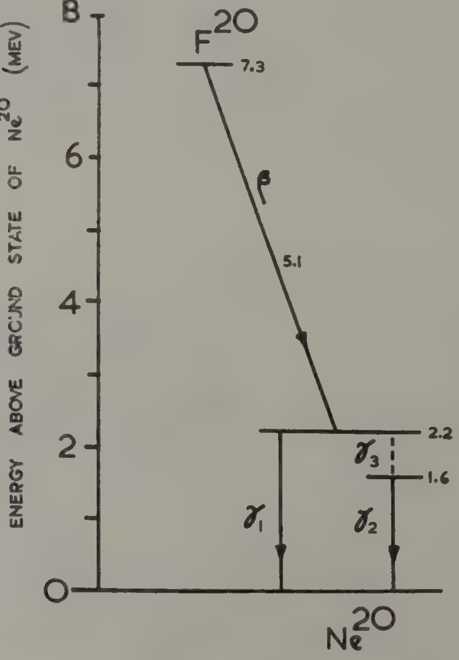
---

* Communicated by Dr. W. E. Burcham.

† Now at the Atomic Energy Research Establishment, Harwell.

This work was completed in the summer of 1949, and a preliminary account of it has already appeared (Jelley 1950). It has recently been repeated in part and extended by R. M. Littauer in the Cavendish Laboratory; his results are presented in the following paper (II.).

Fig. 1.



Accepted decay scheme of F^{20} (decay scheme I).

§ 2. Experimental Technique.

(a) *Production of* F^{20}.

F^{20} was produced by irradiating thick barium or calcium fluoride targets with deuteron beams of 1–50 μamps. at bombarding energies between 500 and 930 keV. The beam was interrupted periodically by an electromagnetically operated shutter to give suitable irradiation and decay periods T_i and T_{ii}. During the irradiation periods, the Geiger counters used in measuring and monitoring the beta- and gamma-radiations were electrically suppressed to avoid high counting rates and to eliminate counting of "prompt" radiations from the target.

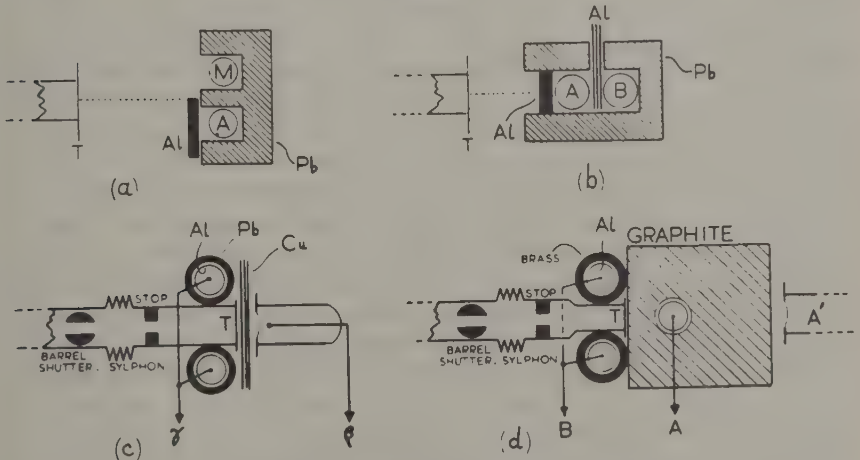
The essential features of the layout of the shutter box, analysing magnet and target tube have been described elsewhere (Jelley and Paul 1948).

Background counting rates, due to shutter radiations, activities present in the analysing magnet, and target activities of longer period than those due to F^{20}, were obtained by stopping the commutator used to operate the cyclic irradiation procedure, leaving the beam on the shutter and counting after a lapse of time for the F^{20} to decay.

(b) *Beta-ray absorption* (see fig. 2 (*a*)).

Two cylindrical aluminium counters, of diameter 2·2 cm., length 6·0 cm. and wall thickness 0·034 gm./cm.^{2} were mounted in a lead housing as shown, with a partition of thickness 11·6 gm./cm.^{2}. Counter M was used as a monitor while various absorbers were placed in front of counter A.

Fig. 2.



Counter arrangements for study of F^{20} activity.
(*a*) Beta-ray absorption.
(*b*) Gamma-ray absorption.
(*c*) Beta-gamma coincidences.
(*d*) Gamma-gamma coincidences.

The counters were situated 28 cm. from the target, T. In this experiment a beam of 50 μamp. at 750 keV. was used, to irradiate a thick BaF_2 target, and the values for T_i and T_d were 10·7 and 8·8 sec. respectively.

(c) *Gamma-ray absorption* (see fig. 2 (*b*)).

The method of observing the secondary electron absorption (Bothe and Kolhörster 1929) was used to determine the energies of the gamma-rays. The F^{20} gamma-rays from an irradiated BaF_2 target emerged through the 0·64 gm./cm.^{2} brass target support and entered the " thick " aluminium radiator of thickness 3·4 gm./cm.^{2}; the 5 MeV. beta-rays were excluded by the radiator and target support. The secondary electrons produced in the radiator passed through two counters A and B in coincidence, aluminium absorbers being placed between the counters which were of diameter 2·2 cm., active length 6·0 cm., and of wall thickness 34 mg./cm.^{2}. In this experiment $T_i = 10·7$ sec., $T_d = 5·2$ sec., and the beam energy and current were 900 keV. and 50 μamp. respectively.

(d) *Beta-gamma and gamma–gamma coincidence experiments.*

The arrangement of counters in the beta-gamma coincidence experiments is shown in fig. 2 (*c*). A CaF_2 target was deposited on a 0·68 gm./cm.^{2} brass support and bombarded by a collimated deuteron beam so that the active F^{20} covered an area 4 × 7 mm.^{2}. To provide increased efficiency and to compensate for fluctuations in beam position, the gamma-rays were detected by two counters connected in parallel. These counters were similar to those used in the absorption experiments, but were surrounded by cylindrical lead converters of thickness (2·9 gm./cm.^{2}) adequate to exclude the 5 MeV. beta-rays while providing the highest efficiency for detection of the gamma-rays. The beta-ray counter was a standard cylindrical brass counter with a 10 mg./cm.^{2} mica window. In this experiment copper absorbers were used so that a large solid angle for the beta-rays could be obtained with large absorption thickness. The F^{20} source activity was chosen to give suitable counting rates; it could be varied by altering the beam current or energy.

Fig. 2 (*d*) shows the arrangement of counters used in the gamma–gamma coincidence experiment. As for the beta–gamma experiment a pair of counters connected in parallel were used for one channel of the coincidence arrangement. These (B) were two aluminium counters of the same design as described above, surrounded by brass converters of wall thickness 3·9 gm./cm.^{2} and mounted one each side the glass target tube. This wall thickness, together with the glass tube, was sufficient to eliminate the 5 MeV. beta-rays. The other counter (A) was of special construction, the reasons for which will be discussed later. It consisted of another standard aluminium counter immersed in a graphite converter; the thickness of graphite between the target and counter wall was sufficient to exclude the 5 MeV. beta-rays. A BaF_2 target was used, mounted on a brass support of thickness 0·26 gm./cm.^{2}, the bombarded area of the target was 4 × 5 mm.^{2}. In both these experiments T_i and T_d were set to 1·0 and 2·6 sec. respectively. For the second part of the gamma–gamma coincidence experiment, in which a beta-counter A′ replaces the gamma-ray counter, the counter used in the beta-ray absorption experiment was employed.

Errors resulting from variations in mean source strength during an experiment were minimized by maintaining the single counting rates as nearly constant as possible during a run. Corrections were applied to the observed coincidence rates to allow for the fact that the counting periods were comparable with the half-life of the activity. The coincidence unit was similar to that described elsewhere (Jelley 1948); the coincidence resolving time was checked from time to time and had the average value $0·63 \pm 0·02$ μsec.

(e) *Spectrometer.*

The spectrometer used for the beta- and gamma-ray energy measurements was an iron-cored instrument with 90° focusing, built by R. M. Littauer and described in the following paper (II.). For the beta-ray measurements the spectrometer was connected directly to the target chamber in which the F^{20} was formed. The overall resolving power and geometrical aperture were 3·5 per cent and 0·031 steradians respectively and the instrument was calibrated for the following energies: (i) 148 keV., the energy of the internally converted gamma-ray F-line from ThB; (ii) 6·8 MeV. which is the maximum energy of the Compton electron distribution from the 6·98 MeV. component of the gamma-radiation from the reaction $F^{19}(p\alpha', \gamma)O^{16}$ (Walker and McDaniel 1948). The calibration for gamma-ray energy measurements is described in II.

§ 3. Results.

(a) *Beta-ray energy.*

Fig. 3 shows the absorption curve in aluminium obtained experimentally with the counter arrangement of fig. 2 (*a*).

The curve is normalized, after correcting to a constant counting rate in the monitor, to an effective intensity of 100 per cent at zero absorber thickness. The gamma-ray background, followed out further than that included in the figure, was extrapolated back and intercepted the intensity axis at 12 per cent. The true beta-absorption curve after subtraction of the background is also shown. To obtain the true origin of this curve a correction term, of 0·509 gm./cm.^{2} must be added to the absorption figures to include absorptions in (i) the BaF_2 target layer, (ii) the brass target support, (iii) the counter wall and (iv) the air path between target and counters.

The beta-ray end-point energy was derived by the method of analysis developed by Bleuler and Zünti (1946), in which curves are constructed relating values of the maximum energy $E_{\beta\max}$ to absorber thicknesses d_n corresponding to reductions in intensity by factors $1/2^n$ of that with no absorber; the family of curves published by these workers strictly apply only to the case of allowed Fermi spectra from nuclei of Z-value 20. Correcting for the initial absorption mentioned above and using Bleuler and Zünti's curves, values of 4·98, 4·90, 4·92 and 4·86 were obtained for $E_{\beta\max}$, giving a mean value of $4{\cdot}91 \pm 0{\cdot}08$ MeV. A correction term of +1 per cent must be added in transforming from $Z=20$ to $Z=9$ and the corrected value for $E_{\beta\max}$ is then $4{\cdot}96 \pm 0{\cdot}08$ MeV.

The relatively small spread among the values obtained suggests that the beta-spectrum consists at least predominantly of one component.

Fig. 4 shows the Kurie plot obtained with the spectrometer; the function $\sqrt{[N(E)/f]}$ is plotted against W, where $f=\sqrt{[W^2-1]}$ and $W=[1+(E/m_0c^2)]$, the total electron energy expressed in units of m_0c^2. N(E) is the relative counting rate per unit energy interval. From the intercept at

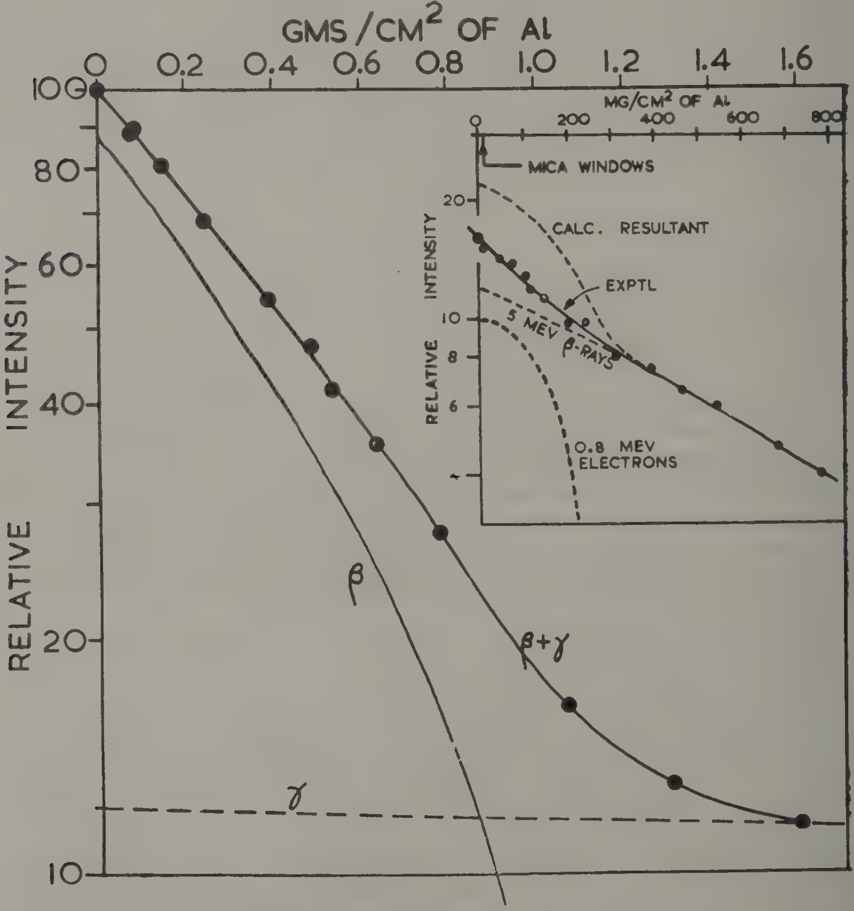
$$W=(10{\cdot}85 \pm 0{\cdot}10)m_0c^2$$

the beta-ray end-point energy was found to be

$$E_{\beta\max}=5{\cdot}03\pm0{\cdot}05 \text{ MeV.}$$

which agrees with that obtained by the absorption method within the limits of error. The beta spectrum also appears from the Kurie plot to consist predominantly of one component.*

Fig. 3.



F^{20} beta-radiation absorption curve.
Inset : more detailed survey of low absorption region.

The period of the beta activity was found to be 12·6±0·1 seconds which is rather higher than the value of 10·7±0·2 sec. recently reported (Snowdon 1950).

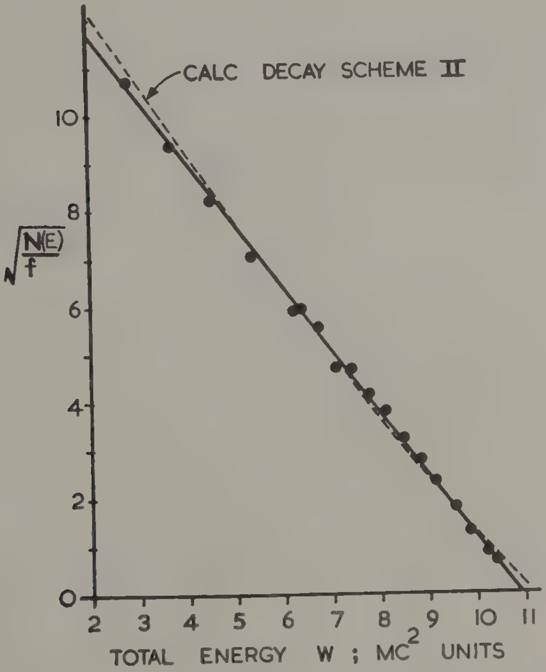
---

* In the more recent experiments reported in II., the presence of a weak beta component of higher energy was established. A rather higher upper limit (5·3 MeV.) for the main beta spectrum was obtained.

(b) *Gamma-ray energy and intensity.*

Fig. 5 shows the secondary electron absorption curve obtained with the counter arrangement shown in fig. 2 (*b*). The ordinates give relative values of the ratio of coincidences to front counter rates (N_C/N_A) normalized to 100 per cent at zero absorber thickness ; the background was obtained from observations with absorber thicknesses greater than those shown in the figure.

Fig. 4.



F^{20} beta spectrum. Kurie plot.

It appears from this absorption curve that there are two main components, with an intensity ratio (by inspection) of between 5 and 10 to 1. The energies of these components were derived by the method due to Bleuler and Zünti (1946, fig. 8), assuming that particular intensity ratio (6·6,1), for the two components which gave the minimum spread in the energy values found for the two radiations. The energies so obtained were :

$E_{\gamma 1}$ 1·87, 1·75, 1·68, 1·62, 1·56, 1·51 and 1·47 MeV.

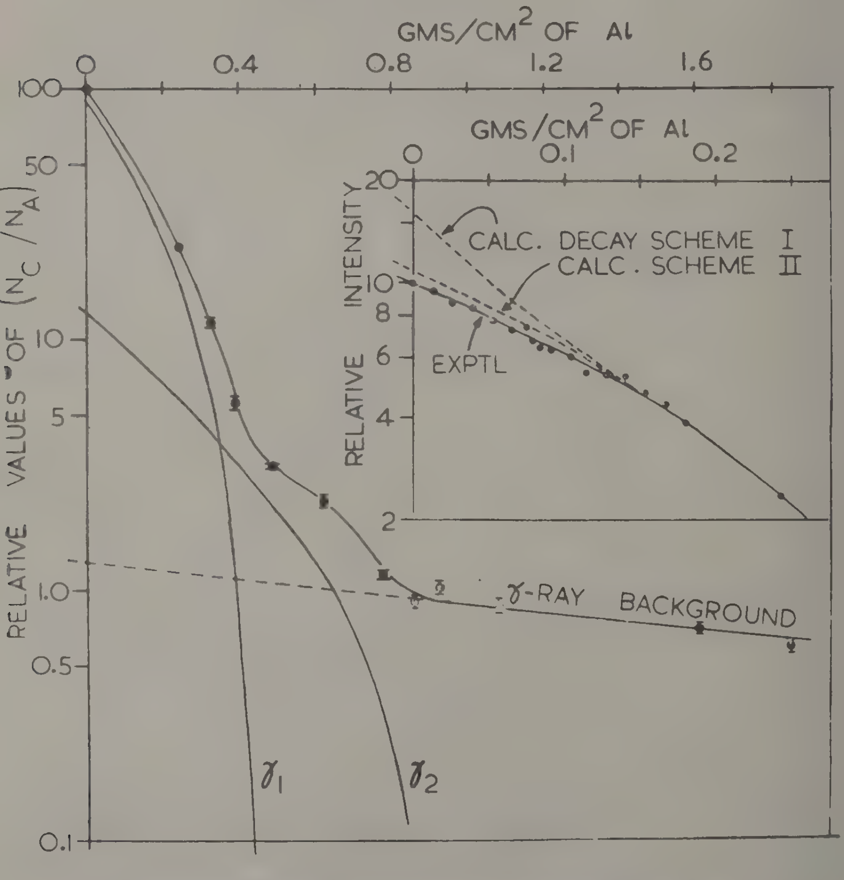
$E_{\gamma 2}$ 2·66, 2·54, 2·41, 2·36, 2·44 and 2·30 MeV.

giving mean values

$$E_{\gamma 1}=1{\cdot}64\pm0{\cdot}05 \text{ MeV.}, \qquad E_{\gamma 2}=2{\cdot}45\pm0{\cdot}06 \text{ MeV.}$$

It was found by experiments with gamma-ray sources of known energy that the ratio of coincidences to front counter rate with no absorber, denoted by $(N_C/N_A)_0$, depended on the incident gamma-ray energy. The ratio of intensities of the two components which gave the best results in the Bleuler–Zünti analysis must therefore be corrected for the energy dependence of the ratio $(N_C/N_A)_0$ to give the true intensity ratio.

Fig. 5.



F²⁰ gamma-radiation. Absorption of secondary electrons.
Inset : more detailed examination of low absorption region.

With this correction, and estimating the errors in extrapolation, it was found that

$$\frac{I_{\gamma 1}}{I_{\gamma 2}} = (8{\cdot}4 \pm 2{\cdot}6)/1,$$

which means relative intensities of 89 per cent and 11 per cent for the soft and hard components respectively.

The results obtained using the spectrometer are described in II. The energy of the hard component could not be measured owing to its low intensity, but that of the softer radiation was found to be $1{\cdot}63 \pm 0{\cdot}02$ MeV., in good agreement with the absorption measurement. The period of the gamma-ray activity was found to be about 12 seconds; no determination of the periods was made for the two components separately (*cf.* II.).

(c) *Possible existence of a lower energy gamma-ray.*

According to the decay scheme I., fig. 1, there should be a gamma-ray of energy 800 keV. (using present values for the other gamma-ray energies) in cascade with the 1·64 MeV. radiation and of equal intensity. In an attempt to establish the existence of this radiation the low energy region of the curve shown in fig. 5 was studied in detail (see fig. 5 inset) using thin absorbers. Fig. 5 (inset) gives, together with the experimental curve, and fitted to it at the point for 0·15 gm./cm.^{2}, the calculated curve for decay scheme I., fig. 1. Comparison of these two curves shows that there is no 800 keV. gamma-radiation present in intensity comparable with that of the 1·64 MeV. component.

The possibility of total internal conversion of the 800 keV. radiation* was investigated by observing the beta-ray absorption curve carefully over the region 0–800 mgm./cm.^{2} of aluminium. Thin windows were used on the target chamber and counters. The curve obtained is shown inset in fig. 3, together with the curve calculated assuming that each quantum of 1·64 MeV. radiation is accompanied by a 800 keV. internal conversion electron. The curves given by Bleuler and Zünti were used for the beta-ray absorption and the data published by Marshall and Ward (1937) for the homogeneous conversion electrons; the calculated and experimental curves are fitted at the 400 mgm./cm.^{2} point. Comparison of the two curves shows that there are no internal conversion electrons of 800 keV. in the right intensity; the slight concave rise at low absorption occurred with other beta-ray sources.

(d) *Beta–gamma coincidences.*

Curran and Strothers (1940) found coincidences between beta- and gamma-rays in the decay of F^{20} and and showed that the absorption of beta-gamma coincidences followed closely that of the beta-rays. Their conclusion that the dominant beta and gamma-ray components are coupled, was confirmed using the counter arrangement shown in fig. 2(*c*). Observation of the single counter rates N_β, N_γ, and the coincidence rate $N_{\beta\gamma}$ were made; the latter was corrected for chance rate N_c using the relation $N_c = 2\tau N_\beta N_\gamma$. A background correction, obtained by observations with a thick absorber, was made to the observed values of N_β to obtain the true counting rate in the β-counter due to F^{20} beta-rays; this correction became very large near the end-point of the beta-ray absorption curve.

---

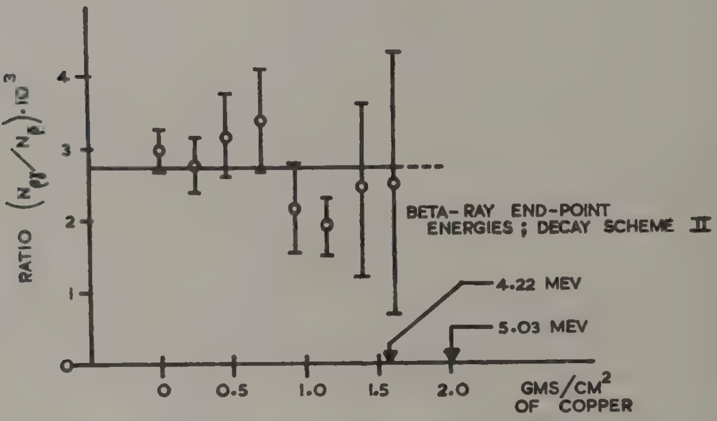
* The author is grateful to Professor S. Devons for suggesting this possibility.

In fig. 6 is shown a plot of the values of the ratio $N_{\beta\gamma}/N_\beta$ versus absorber thickness, the corrections discussed above having been applied. On this figure are also included the cut-off points corresponding to the two beta-ray energies involved in the alternative decay scheme to be presented later. Owing to the poor statistics it was impossible to obtain evidence in favour of a particular decay scheme, though the result was consistent with the previous evidence that at least the dominant beta ray component proceeds to an excited state of Ne^{20}.

(e) *Gamma–gamma coincidences.*

Owing to the high energy of the beta-radiation from F^{20}, some or all of the gamma–gamma coincidences observed by Curran and Strothers (1940) in their arrangement, in which lead-walled counters were used, may have been due to bremsstrahlung. With this in mind it was decided to use the counter arrangement of fig. 2 (*d*), in which one counter was constructed of graphite, in a study of gamma–gamma coincidences.

Fig. 6.



Results of beta-gamma coincidence experiment.

It is essential to carry out a gamma–gamma coincidence experiment on a quantitative basis if the resulting evidence is to be used to assist in the designation of a particular decay scheme. With this aim in view, a method originally proposed by Dunworth (1940) and more recently developed by Wiedenbeck (1947), was applied to a study of gamma–gamma coincidences from F^{20}. Essentially the method consists firstly of carrying out a gamma–gamma coincidence experiment and, secondly, repeating it after one of the gamma-ray counters has been replaced by a beta-ray counter; the single and beta–gamma and gamma–gamma coincidences are measured, from which various ratios can be set up; these ratios, which are different for different decay schemes, are then compared with calculated values, using measured values of counter efficiencies.

With the arrangement already described, sets of gamma–gamma and beta–gamma coincidence runs were carried out, counter A being replaced by the beta-counter A′ for the beta–gamma runs, taking care not to disturb the geometry of the counters B. Ten gamma–gamma and four beta–gamma runs were made. As before, the individual and coincidence rates were found from the number of irradiation cycles and counting periods, after a predetermined number of coincidences had accumulated; these figures were set at 200 and 400 for gamma–gamma and beta–gamma runs respectively.

Correcting the single counter rates for backgrounds and the observed coincidence rates for chance coincidences, the following ratios were obtained after compounding the ten gamma–gamma and four beta–gamma runs :—

$$\left.\begin{array}{ll}\text{From the } \gamma\gamma \text{ runs } N_{\gamma\gamma}/N_{A\gamma} & =(3\cdot03\pm0\cdot23)\times10^{-4}.\\ \text{From the } \beta\gamma \text{ runs } N_{A'\beta}/N_{\beta\gamma} & =523\pm40.\end{array}\right\}$$

The errors quoted here are standard deviations calculated from the spread among the ratios of the varoius rates in the two types of runs.

From the above figures, the value obtained for the ratio R, defined by

$$R=\frac{N_{\gamma\gamma}}{N_{A\gamma}}\cdot\frac{N_{A'\beta}}{N_{\beta\gamma}}$$

is

$$R_{obs}=0\cdot16\pm0\cdot02.$$

R is a function of the gamma-counter efficiences and the branching ratio of the altenative gamma-transitions; inserting the various values for these quantities, the derivations of which are discussed in Appendix A, the value of R for decay scheme I (fig. 1) is found to be

$$R_{calc}=0\cdot47.$$

It will be seen that the experiment gives too low a gamma–gamma coincidence rate for decay scheme I. and, furthermore, it is likely that some or all of these observed coincidences are false since :

(*a*) an approximate calculation of the effects of bremsstrahlung showed that even with the graphite counter, enough coincidences could exist to account wholly for the observed rate ;

(*b*) the formation of N^{13} on the target, due to the reaction $C^{12}(dn)N^{13}$ could have provided a source of coincident radiations comparable in strength with that observed.

A repetition of this experiment under better conditions (II.) has confirmed that the number of genuine gamma–gamma coincidences in the decay of F^{20} is very small.

§ 4. DISCUSSION.

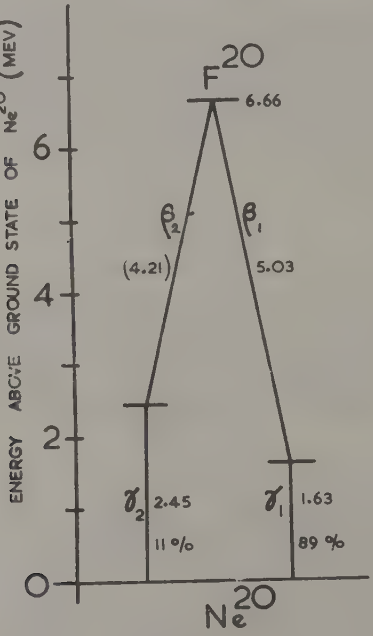
(a) *The Decay Scheme of* F^{20}.

The small number of genuine gamma–gamma coincidences and the absence of a gamma-ray of energy 800 keV. in intensity comparable with that of the 1·64 MeV. radiation from F^{20} are incompatible with decay scheme I. (fig. 1). The results are, however, consistent with the decay scheme II., fig. 7, in which the 1·64 MeV. gamma radiation follows the beta particle emission. The fitting of the 2·45 MeV. gamma-radiation into this scheme requires a complex beta particle spectrum, and a slightly curved Kurie plot should have been found. The expected deviation from linearity (shown in fig. 4) is, however, small and would have escaped detection.

(b) *The mass of* F^{20}

Direction determinations of the mass of F^{20} in terms of that of F^{19} can be obtained from the energy release in the reactions $\mathrm{F}^{19}\,(dp)\,\mathrm{F}^{20}$ and $\mathrm{F}^{19}\,(n\gamma)\,\mathrm{F}^{20}$. A repetition of the observations of Bower and Burcham (1939) on the $\mathrm{F}^{19}\,(dp)$ reaction was made (see Appendix B), and a Q-value of $4{\cdot}48 \pm 0{\cdot}04$ MeV. was obtained. Using the masses of the proton, deuteron and alpha particle given by Rosenfeld (1948), and for the mass of F^{19} the value $19{\cdot}00449 \pm 0{\cdot}00003$ * (Tollestrup, Fowler and Lauritsen 1950) the mass of F^{20} is found to be $20{\cdot}00626 \pm 0{\cdot}00005$. The observations of Kinsey, Bartholomew and Walker (1950) on the energy of the capture radiation in the $\mathrm{F}^{19}\,(n\gamma)$ reaction lead to a mass for F^{20} of $20{\cdot}00632 \pm 0{\cdot}00006$. The average of these two values is $20{\cdot}00629 \pm 0{\cdot}00005$.

Fig. 7.



Proposed decay scheme of F^{20} (decay scheme II.).

---

* This recently published value has been adopted in preference to that used in the first report of these results (Jelley 1950).

The present results enable the mass of F^{20} to be found in terms of that of Ne^{20}, for which the value 19·99890±0·00006 (Rosenfeld 1948) is assumed. Using the energy values $E_\beta=5\cdot03\pm0\cdot05$ MeV. and $E_\gamma=1\cdot63\pm0\cdot02$ MeV. the energy release in the decay of F^{20} is, according to decay scheme II. (fig. 7) 6·66±0·05 MeV. and the mass of F^{20} becomes 20·00606±0·00008, which is rather lower than the previous estimates. If this value is correct, an additional long-range proton group might be expected in the F^{19} (*dp*) reaction, but this could not be found. It seems possible, however, from the work reported in II. that the beta-particle energy may have been underestimated in the present experiments.

APPENDIX A.

Determination of the gamma-ray counter efficiencies and the evaluation of R *in the gamma-gamma coincidence experiment.*

Before evaluating the quantity R defined in §3 (e), it was necessary to determine the gamma-ray counter efficiencies at a number of gamma-ray energies. We will here denote these efficiencies by ϵ with two suffixes, the first indicating the energy of the radiation and the second to denote which counters are meant (*cf.* fig. 2(*d*)). The calibration was carried out by placing a 1·4 μC Co^{60} source in the centre of the target area and finding the single and coincidence counting rates with the counters in their normal position. Assuming the mean energy of the gamma-radiation to be 1·2 MeV. and a zero angular correlation between the two components, the following absolute values for the overall efficiencies of the two counters were obtained

$$\epsilon_{1\cdot2\mathrm{A}}=7\cdot3\times10^{-4} \text{ and } \epsilon_{1\cdot2\mathrm{B}}=2\cdot0\times10^{-3}.$$

Use was then made of curves published recently by Maier-Leibnitz (1946) to obtain the relative values of the efficiencies at other energies ; in this way the absolute efficiencies were determined for the following energies :

Counter	0·8	1·6	2·4	MeV.
A (graphite+Al)	4·8	9·4	13·0	$\times10^{-4}$
B (brass+Al)	1·4	2·7	4·1	$\times10^{-3}$

In this calculation it was assumed that counter A was made entirely of aluminium and counters B of brass ; the corresponding curves given by Maier-Leibnitz being employed. These assumptions are justified since (*a*) the aluminium walls of the counters proper were of a thickness small compared to the range of the secondary electrons, (*b*) graphite and aluminium have comparable values of density and atomic number, and (*c*) over the range of energies considered the form of the efficiency-vs. energy curve is similar for aluminium and brass.

The ratio R, defined in § 3 (e) is a function of these efficiencies and the relative intensities of the radiations. The value of R for decay scheme I. assuming no angular correlation between the various gamma-ray components, is given by :

$$R_{\text{calc}}=\frac{f(\epsilon_{1\cdot 6B}\cdot\epsilon_{0\cdot 8A}+\epsilon_{0\cdot 8B}\cdot\epsilon_{1\cdot 6A})}{[f(\epsilon_{1\cdot 6B}+\epsilon_{0\cdot 8B})+(1-f)\epsilon_{2\cdot 4B}][f(\epsilon_{1\cdot 6A}+\epsilon_{0\cdot 8A})+(1-f)\epsilon_{2\cdot 4A}]}$$

where f and $(1-f)$ are the relative intensities of the (1·6 ; 0·8) and 2·4 MeV. transitions respectively. Inserting the absolute efficiencies obtained as described above and the value for f from § 3 (b) a value of 0·47 is obtained for R_{calc}. For decay scheme II $R_{\text{calc}}.=0$.

APPENDIX B.

Determination of the Q-value of the Reaction F^{19} (*dp*) F^{20}.

In this experiment a thin BaF_2 target was bombarded by deuterons and the proton groups from the above reaction detected in a proportional counter. Calibrated mica absorbers were used to vary the residual range so that the various groups could be studied separately. Another proportional counter with a fixed amount of absorber was used for monitoring purposes. Direct resolution of the groups was found possible using a discriminator operating at high bias.

The protons of maximum energy were observed at three different angles with respect to the deuteron beam, 30°, 90° and 150° and their range was obtained by direct comparison with the range of the protons from the reactions C^{13}(*dp*) C^{14} and Be (*dp*) Be^{10} for which accurate Q-values are available. Throughout these experiments a deuteron energy of 930±15 keV. was used. The following Q-values were obtained :

Reference Reaction	Angle of Observation	Q MeV.
Be (*dp*) Be^{10}	30°	4·41
Be (*dp*) Be^{10}	90°	4·40
Be (*dp*) Be^{10}	150°	4·44
C^{13} (*dp*) C^{14}	30°	4·67
C^{13} (*dp*) C^{14}	90°	4·46
C^{13} (*dp*) C^{14}	150°	4·52

giving a mean value of 4·48±0·04 MeV.

ACKNOWLEDGMENTS.

The Author wishes to thank Dr. W. E. Burcham for his continual interest and advice through the work. Thanks are also due to R. M. Littauer for making available his spectrometer, to C. D. Curling and J. O. Newton who lent their apparatus for the (*dp*) experiment, and to W. Birtwhistle for maintenance and operation of the high-tension equipment.

The Author is indebted to the Department of Scientific and Industrial Research for financial assistance during the period of the work.

REFERENCES.

BLEULER, E., and ZUNTI, W., 1946, *Helv. Phys. Acta.*, **19**, 375.
BONNER, T. W., 1939, *Phys. Rev.*, **56**, 207.
BOTHE, W., and KOLHÖRSTER, W., 1929, *Z. f. Physik*, **56**, 751.
BOWER, J. C., and BURCHAM, W. E., 1939, *Proc. Roy. Soc.*, A, **173**, 379.
CURRAN, S. C., and STROTHERS, J. E., 1940, *Proc. Camb. Phil. Soc.*, **36**, 252.
DUNWORTH, J. V., 1940, *Rev. Sci. Instrum.*, **11**, 167.
HORNYAK, W. F., and LAURITSEN, T., 1948, *Rev. Mod. Phys.*, **20**, 191.
JELLEY, J. V., 1948, *Canad. J. Res.*, **26**A, 255 ; 1950, *Proc. Phys. Soc.*, A, **63**, 538.
JELLEY, J. V., and PAUL, E. B., 1948, *Proc. Camb. Phil. Soc.*, **44**, 133.
KINSEY, B. B., BARTHOLOMEW, G. A., and WALKER, W. H., 1950, *Phys. Rev.*, **78**, 481.
MAIER-LEIBUITZ, H., 1946, *Zeits. f. Naturforschung*, **1** (5), 243.
MARSHALL, J. S., and WARD, A. G., 1937, *Canad. J. Res.*, **15**A, 39.
MATTAUCH, J., and FLAMMERSFELD, A., 1949, *Isotopenbericht Zeits. f. Naturforschung.*
POWELL, C. F., 1940, *Nature, Lond.*, **145**, 155.
ROSENFELD, L., Nuclear Forces Fasc. II (Amsterdam : North Holland Publishing Co.).
SNOWDON, S. C., 1950, *Phys. Rev.*, **78**, 299.
TOLLESTRUP, A., FOWLER, W. A., and LAURITSEN, C. C., 1950, *Phys. Rev.*, **78**, 372.
WALKER, R. L., and MCDANIEL, B. D., 1948, *Phys. Rev.*, **74**, 315.
WIEDENBECK, M. L., 1947, *Phys. Rev.*, **72**, 429.

CVIII. *The Radioactivity of* F^{20}.—II.

By R. M. LITTAUER*,
Cavendish Laboratory, Cambridge †.

[Received August 23, 1950.]

SUMMARY.

Experiments on the radioactivity of F^{20} are described which give the following results : 3·5 per cent of the disintegrations go direct to the ground state of Ne^{20}, while the remainder leave Ne^{20} in an excited state of energy $1{\cdot}63 \pm 0{\cdot}02$ MeV. The best present estimate of the F^{20} mass is $20{\cdot}006\ 31 \pm 0{\cdot}000\ 05$, corresponding to a total energy release $F^{20}-Ne^{20}$ of $6{\cdot}87 \pm 0{\cdot}07$ MeV. A weak gamma ray of energy $1{\cdot}0 \pm 0{\cdot}1$ MeV. is also observed, but its allocation to the decay of F^{20} is not established. Gamma–gamma coincidence experiments limit the fraction of disintegrations which produce *cascade* gamma rays to less than 2 per cent, unless one of the rays has a life-time longer than 10^{-6} sec.

§ 1. INTRODUCTION.

IN the preceding paper, hereafter referred to as I., J. V. Jelley has reported some experiments concerned with the short-lived beta emitter F^{20}. The main conclusion to be drawn from his work is the that dominant beta spectrum is linked with a single gamma ray. This result is borne out in the present investigation, which was undertaken with the object of clarifying some of the doubtful points raised in I. and of obtaining a more precise estimate for the mass-value of F^{20}.

§ 2. EXPERIMENTAL METHOD.

(a) *Production of* F^{20}.

F^{20} was produced throughout this work by intermittent bombardment of thin CaF_2 targets with 930 keV. deuterons, as described in I. The radiations were studied by spectrometer and coincidence counting methods. In the spectrometer experiments, the irradiation and counting periods were approximately equal in order to achieve maximum statistical accuracy. In the coincidence work, where the useful source strength is in any case limited, the bombarding periods were shortened and a greater fraction of the cycle was available for counting. The period of the switching operations was always short compared with the life-time of the decay.

---

* Communicated by Dr. W. E. Burcham.

† Now at Cornell University, Ithaca, N.Y.

(b) *The Beta Spectrometer.*

The spectrometer is a compact instrument of the 90° magnetic focusing type. This is a special case of the general group discussed by Stephens (1934): any homogeneous field extending over a wedge-shaped area gives simple refocusing of charged particles originating from a line source. It can easily be shown that the resolving power of such a spectrometer for a given transmission factor is independent of the angle of the wedge. The 90° instrument has the advantage over the more conventional 180° type that the source and detector lie outside the field gap, and that their separation is greater, which facilitates shielding. It has the disadvantage that the stray field at the edges of the pole-pieces modifies the trajectories of the particles.

This effect was not considered in the design of the instrument used for the work on the beta particle spectrum reported in I.; there was, moreover, some uncertainty in the calibration of this instrument, so that the values obtained with it must be regarded as only approximate.

For the design of the present spectrometer, the stray field was mapped with a spinning-coil fluxmeter, and a simple graphical method was used to plot the trajectories of electrons on this map. In this way it was easy to determine the position of source and detector for optimum focusing in the presence of the stray field.

Fig. 1 shows a cross-section in the median plane of the spectrometer (horizontal). For measurement of the beta spectra an internal target arrangement was used, as shown at (*a*). The active layer was deposited by evaporation *in vacuo* on a thin (0·5 mm.) copper backing (cooled at one side by water), which was mounted at 45° and had a projected width of 4 mm. The experimental resolving power with a source of this width is 1·8 per cent; the solid angle of acceptance of the spectrometer is 0·6 per cent of 4π.

For monitoring the activity of F^{20} produced on the target, the energetic beta rays emerged through the backing and the window W, and were counted in a pair of coincidence counters not shown in the diagram. The spectrometer counters were also operated in coincidence to reduce the background due to stray gamma radiation; the shielding from the target was completed by building up lead blocks outside the vacuum box.

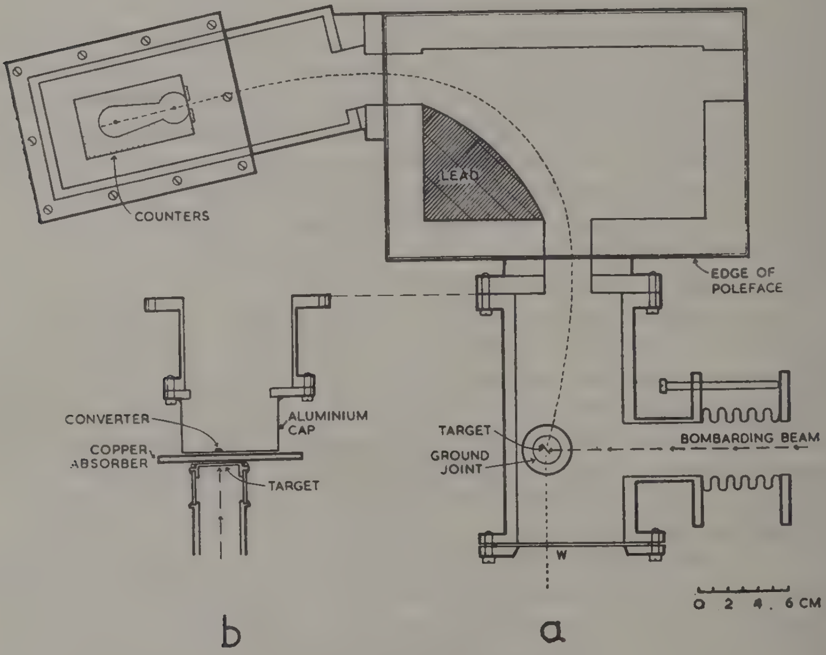
For measuring gamma ray energies, the target arrangements shown in fig. 1 (*b*) were employed. To stop the energetic beta rays from entering the spectrometer, a sheet of copper (3·3 g/cm.^{2}) was interposed. This, and the thin aluminium wall of the spectrometer, acted as converter for Compton electrons; no localization of this converter was possible apart from the fact that the bombarding beam was relatively well focused and thus restricted the area of material which was effective. These conditions were of course reproduced during the calibration of the spectrometer for Compton distributions.

For an examination of photo-electrons, a thin (50 mg./cm.^{2}) lead radiator, 4 mm. wide, was mounted on the inside of the alumium cap.

The target activity was monitored by means of its gamma rays, which were recorded in an arrangement of coincidence counters. All readings were normalized to this monitor.

The spectrometer was calibrated by reference to the following well-known energies: the F line of Thorium B at 148 keV., the gamma rays of Co^{60} at 1·172 and 1·322 MeV. (Lind, Brown and DuMond 1949), and the ThC″ line at 2·62 MeV. Since the iron of the magnet is far from saturation and the main part of the reluctance of the magnetic circuit lies in the air gap, the field is very accurately proportional to the magnet current up to electron energies of 7 MeV. The fact that the calibration could be extrapolated to this region was verified by examining the gamma rays from $F^{19}(p\alpha', \gamma)O^{16}$ (at 6·13 and 6·98 MeV.; Walker and McDaniel 1948) with the spectrometer.

Fig. 1.



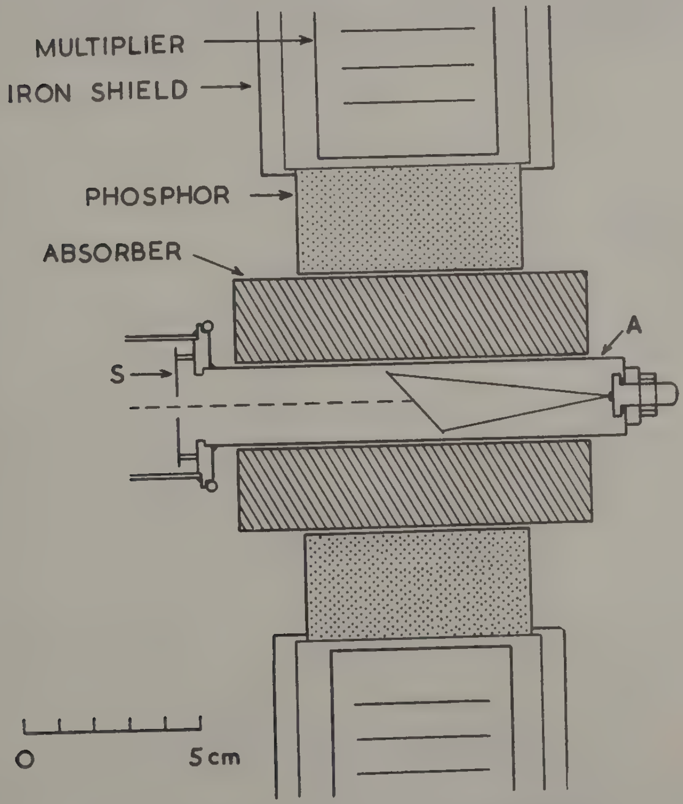
Cross-section of spectrometer.

The magnet current was stabilized by an electronic stabilizer using four valves of Mazda type 12E1 as control element. Relative values were read from a secondary standard meter, accurate to 0·2 per cent.

(c) *Coincidence Experiments.*

Fig. 2 shows the target arrangements used for the coincidence experiments. The target was deposited on a thin molybdenum backing supported by two wires in the aluminium can A. (This can was lined with very thin brass foil to avoid interaction of the bombarding deuterons with the aluminium.) The beam restricting aperture S was also of molybdenum, uncooled. Both the target and S reached dull red heat during a bombarding cycle; this prevented the formation of carbon from decomposing pump vapours, and thus avoided the production of N^{13} by the reaction $C^{12}(d, n)$. The N^{13} provided a troublesome background of genuine gamma-gamma coincidences in the earlier parts of the work, as has now been verified by following the decay of these coincidences (half-life 10 mins.).

Fig. 2.



Target arrangements for coincidence experiments.

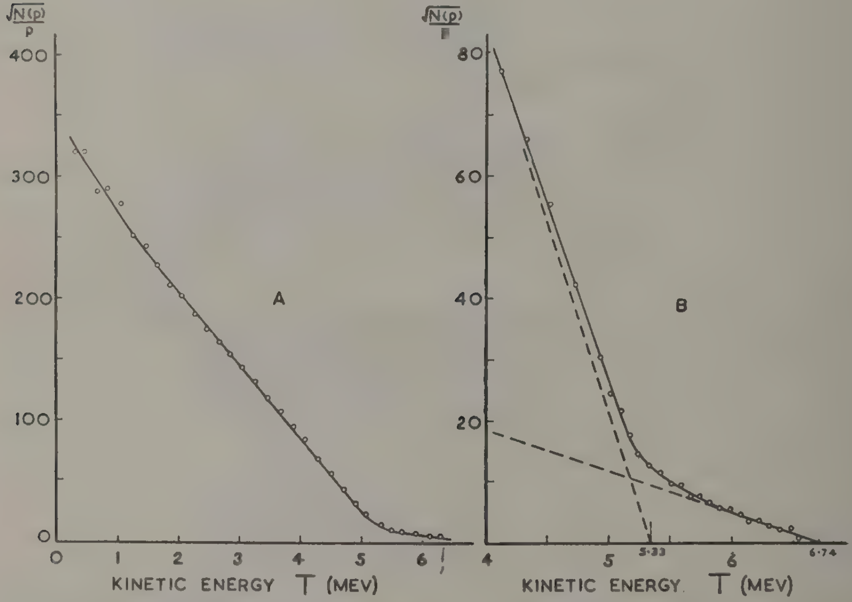
Scintillation counters were used to detect the gamma rays, as they are more efficient than the Geiger counters used in I., and permit higher counting rates to be used. These counters are shown in the 180° position in fig. 2; to allow for a possible strong angular correlation between the two quanta, duplicate experiments with the counters at right angles were also made, but yielded no different results. An additional improvement made in this work was the continuous recording of the chance coincidence rate by a method described elsewhere (Littauer 1950). This made it possible to use sources of such a strength that the chance rate exceeded the genuine rate, and also facilitated an immediate interpretation of the results. The coincidence resolving time was 10^{-6} sec.

§ 3. Results.

(a) *The Beta Particle Spectrum.*

Fig. 3 shows a Kurie plot, constructed in the usual way, of the observed spectrum from F^{20}. It is obvious that the spectrum is complex, with a main component of end-point $5{\cdot}33 \pm 0{\cdot}05$ MeV., and a weaker one (3·5 per cent) of end-point $6{\cdot}74 \pm 0{\cdot}1$ MeV.* There is no evidence of the presence of any internal conversion line in the region 0·15–7 MeV.

Fig. 3.



(A) Kurie plot from F^{20} beta ray spectrum.
(B) Detailed examination of high-energy region.

(b) *The Gamma Ray Spectrum.*

The results of an examination of the Compton recoil electrons are exhibited in fig. 4. The experimentally determined shape due to a monochromatic source is shown fitted to this plot, and it is clear that the points in the low energy region depart significantly from this shape.

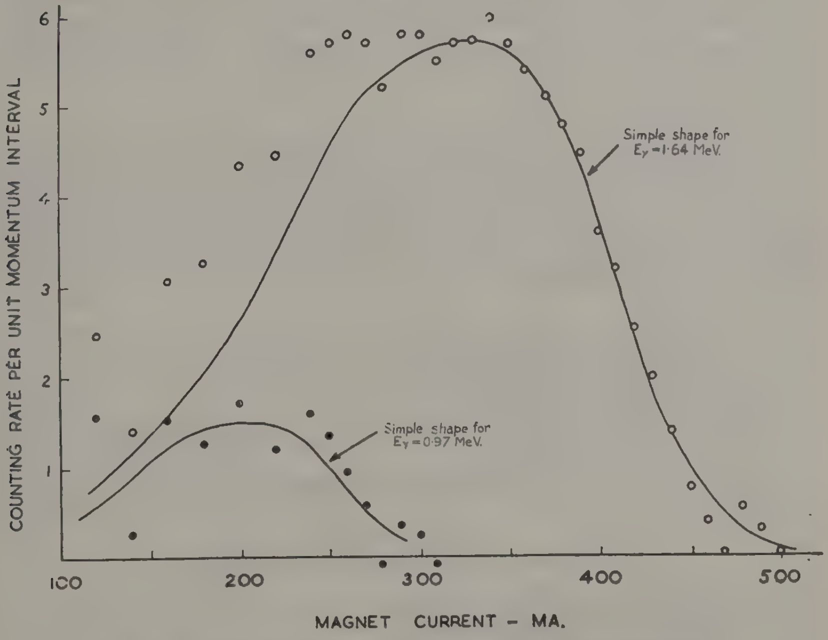
---

* The energy of the main component differs appreciably from that obtained with the original, uncorrected spectrometer (*cf.* I.); the present work is regarded as more reliable. The earlier work did not reveal the higher energy component.

The observed differences are plotted on the same graph, and can be fitted reasonably well by another simple shape. It should be pointed out that the yields in this experiment were very low indeed and that the background, subtracted in Fig. 4, represented about 20 per cent of the maximum counting rate. The two gamma lines indicated have energies 1·64±0·03 and 1·0±0·1 MeV., and intensities roughly as 4:1. Other gamma lines in the region 0·7–3 MeV. would have been detected if their intensity had exceeded about 10 per cent of that of the main component.

An attempt to detect the photo-electrons from the main gamma component met with only fair success. The line appeared to be indicated but the intensity in this experiment was not sufficiently high to yield quite definite results. The gamma energy deduced from the position of the line is 1·62±0·02 MeV., in good agreement with the value derived from the Compton distribution. The energy 1·63±0·02 MeV. has been adopted as the best estimate.

Fig. 4.



Compton electron distribution from the gamma rays of F^{20}.

The intensity obtainable with the spectrometer did not allow the high-energy gamma ray reported in I to be observed. Better efficiency for gamma ray detection can be achieved with an ionization chamber using liquid argon, in which the Compton secondaries give up all their energy. G. W. Hutchinson (1948) has developed such a "liquid argon spectrometer" in the Cavendish Laboratory; at the request of the author, a study of the F^{20} gamma ray spectrum was made with this instrument.

The intensity was sufficiently high to allow an examination of the lifetimes of the individual gamma components, whereas in the work reported in I. it was possible to check only that the average half-life of the gamma activity was consistent with its allocation to F^{20}.

This investigation * revealed the presence of two gamma ray lines of energy 1·64 and 1·0 MeV. and relative intensity 4 : 1 approximately, in agreement with the spectrometer work. A complex component of energy up to 4 MeV., not easily resolvable into separate lines, was also found. The decay periods of the 1 MeV. line and of the high-energy component were compared with that of the main line at 1·64 MeV., which is known to belong to the F^{20} decay (*cf.* I). The evidence concerning the 1 MeV. line was not conclusive, but did not contradict its allocation to the decay of F^{20}. The period of the higher energy component, however, was definitely longer (∼1 min.), and it is therefore likely that this component, which may almost certainly be identified with the 2·45 MeV. radiation reported in I., does not in fact arise from the decay of F^{20}, but may be attributed to an unidentified contamination.

(c) *Gamma–gamma coincidences.*

In the analysis of the results of these experiments, we have assumed the efficiency of detection for gamma rays to be approximately independent of energy in the region 0·6–2·0 MeV., which introduces no error greater than about 30 per cent when the scintillation counters are used at low settings of discriminator bias.

If we put

α, β=efficiencies in the two gamma channels,
τ=coincidence resolving time,
$\bar{n}$=mean source strength,
p=fraction of disintegrations yielding two gamma rays in cascade,
T=time of counting,

we then have : the single counts recorded in channel A (after subtraction of the background)

$$S(A)=2p\bar{n}\alpha T+(1-p)\bar{n}\alpha T=(1+p)\bar{n}\alpha T.$$

Similarly $S(B)=(1+p)\bar{n}\beta T$.

The number of randon coincidences is

$$R=2(1+p)^2\alpha\beta\tau \,.\, \overline{n^2} \,.\, T,$$

and the genuine coincidences are given by

$$G=2p\bar{n}\alpha\beta T.$$

Hence

$$p/(1+p)=(\tau/2\alpha\beta T)^{\frac{1}{2}} \,.\, G/\sqrt{R}$$

$$=(1/S) \,.\, G/\sqrt{R},$$

where

$$S=(2\alpha\beta T/\tau)^{\frac{1}{2}}.$$

---

* The author is indebted to Mr. Hutchinson for permission to use these results before publication.

(We have assumed $\overline{n^2}=(\overline{n})^2$, which is approximately true for F^{20} under the conditions of the experiment.)

To measure S a source of known p (*e. g.* Co^{60} with $p=1$) was used, placed in the position of the target, with the same absorbers as in the corresponding F^{20} experiment.

Table I. gives the results of the experiments as a function of the absorbing material used to stop the hard beta rays.

TABLE I.

Apparent values of p with various absorbing materials. (p is the fraction of disintegrations which yield cascade gamma rays).

Absorber	G/5 min.	R/5 min.	S	p (%)
Graphite	5 ± 10	230	42	$0\cdot8\pm1\cdot6$
Aluminium	50 ± 20	450	70	$3\cdot4\pm1\cdot4$
Copper	170 ± 40	660	120	$5\cdot8\pm1\cdot3$
Lead	145 ± 40	850	31	19 ± 5

It is evident that the apparent value of p rises with the atomic number of the absorber, and that the true value must be well below 2 per cent. This demonstrates the effect of bremsstrahlung from the hard beta rays, and also shows that there cannot be a cascade of gamma rays (of life-time less than 10^{-6} sec.) in the decay of F^{20} with fractional intensity above 2 per cent.

Rough experiments were also made on beta-gamma coincidences from F^{20}. These established that there is no beta-gamma angular correlation greater than 7 per cent, and that there are $1\cdot06\pm0\cdot1$ gamma quanta per beta particle. It has not been possible to separate the high-energy beta component for this measurement on accouut of the low intensity near the end-point of the spectrum.

§ 4. DISCUSSION.

(*a*) *The Mass of* F^{20}.

An estimate of the mass difference $F^{20}-Ne^{20}$ can be obtained (as in I.) by adding the energies of the prominent gamma line and of the main beta spectrum, for there can be no doubt that these must occur in cascade. Thus the total energy released in the decay is $6\cdot96\pm0\cdot06$ MeV., which corresponds to a mass difference of $7\cdot47\pm0\cdot06$ mMU. On the basis of a Ne^{20} mass of $19\cdot998\ 90\pm0\cdot000\ 06$ (Rosenfeld 1948), this gives a F^{20} mass of $20\cdot006\ 37\pm0\cdot000\ 08$.

Another estimate of the F^{20} mass can be obtained from the mass of F^{19} and the Q of either of the reactions $F^{19}(d, p)$ or $F^{19}(n, \gamma)$. These reactions give a mean mass difference $F^{20}-F^{19}$ of $1\cdot001\ 80\pm0\cdot000\ 04$ (*cf.* I.); if we take the F^{19} mass to be $19\cdot004\ 49\pm0\cdot000\ 03$ (Tollestrup, Fowler and Lauritsen 1950), then the F^{20} mass becomes $20\cdot006\ 29\pm0\cdot000\ 05$.

The two estimates of the F^{20} mass differ by 0·08±0·10 mMU, which indicates a satisfactory agreement. The most likely present estimate of the mass of F^{20} is

$$20{\cdot}006\ 31 \pm 0{\cdot}000\ 05.$$

(b) *The Decay Scheme of* F^{20}.

If the high-energy beta component corresponds to a decay of F^{20} direct to the ground state of Ne^{20}, its maximum energy, on the basis of the present spectrometer measurements, should be 6·96±0·06 MeV. The experimental upper limit of the distribution is 6·74±0·1 MeV., which differs from the expected value by 0·22±0·12 MeV. Despite this small discrepancy, however, it is assumed that the transition does indeed go to the ground state; there is no evidence to support an alternative hypothesis.

Table II. gives the ft. values (*cf.* Konopinski 1943) of the two partial beta components; these values are not inconsistent with the assumption that the transitions to the excited state and ground state of Ne^{20} are respectively allowed (unfavoured) and second forbidden, as required by the nuclear shell model (Feenberg and Hammack 1949). On this assumption, and taking the excited state of Ne^{20} concerned to be a (2, +) state the ground state of F^{20} should be of even parity and spin 2 or 3.

TABLE II.

ft. values for the two partial beta ray spectra.

E_{max}	*t*	*f*(E)	*ft*
11·4 mc^2	12 secs.	$6{\cdot}5\times10^3$	$0{\cdot}8\times10^5$
14·2	350	2×10^4	$0{\cdot}7\times10^7$

Fig. 5 shows a number of decay schemes that might be postulated on the basis of these results. Scheme I, which contains only definitely established transitions, is incomplete in so far as it does not account for the presence of the 1 MeV. gamma radiation; it must be remembered, however, that the allocation of this to F^{20} is still open to question. (The 2·45 MeV. component reported in I. has been omitted on the basis of the present results.)

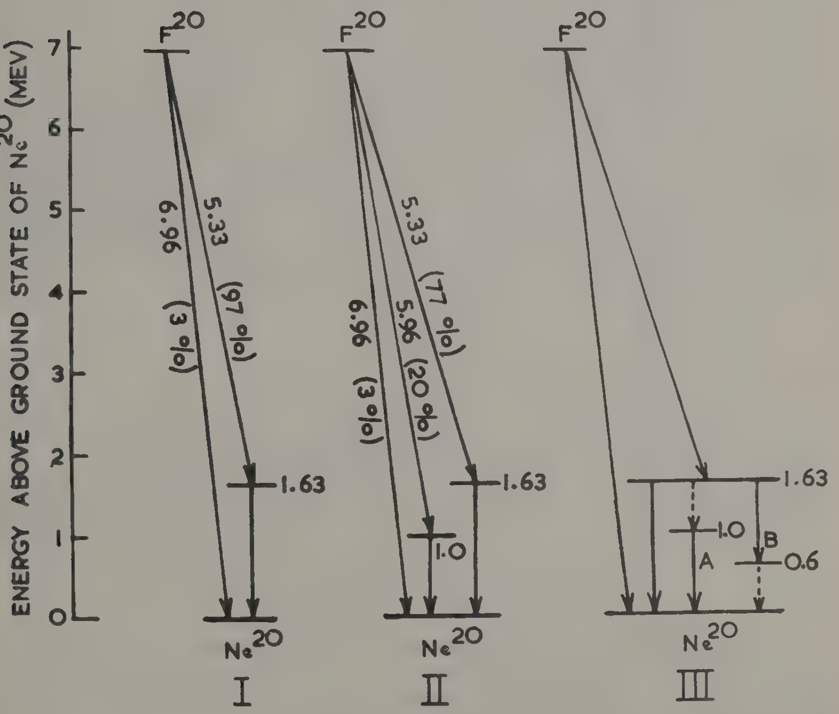
Decay shemes II. and III. are more speculative and would require further evidence to establish them. Scheme II. postulates a beta component of energy about 6·0 MeV., linked with the 1 MeV. gamma ray. We have attempted to analyse the experimental Kurie plot into three components, as shown in fig. 6. The maximum intensity of the 6 MeV. component that can be accommodated is about 5 per cent, which is insufficient for the requirements of scheme II. The broken curve in fig. 6 shows the Kurie plot* that would be expected if scheme II. were indeed correct. It is seen that little evidence to support this scheme can be adduced.

---

* In the construction and analysis of the Kurie plots the effect of the finite spectrometer resolution has been neglected, as this is small compared with the other experimental uncertainties.

Scheme III. (A or B) is consistent with the data from the coincidence experiments only if the lower of the cascade gamma transitions has a life-time longer than 10^{-6} sec., so that the quanta should not be recorded as coincident. If the level arrangement B is correct, so that the gamma transition has ouly 600 keV. of energy available, a spin $\geqslant 3$ for the postulated level would suffice for this requirement. This possibility might repay further study. (A total internal conversion of the second quantum is ruled out by the results of both I. and the present work.)

Fig. 5.


ENERGY ABOVE GROUND STATE OF Ne^{20} (MEV)
0
1
2
3
4
5
6
7
F^{20}
6.96 (3 %)
5.33 (97 %)
1.63
Ne^{20}
I
F^{20}
5.33 (77 %)
6.96 (3 %)
5.96 (20 %)
1.0
1.63
Ne^{20}
II
F^{20}
1.63
1.0
A
B
0.6
Ne^{20}
III

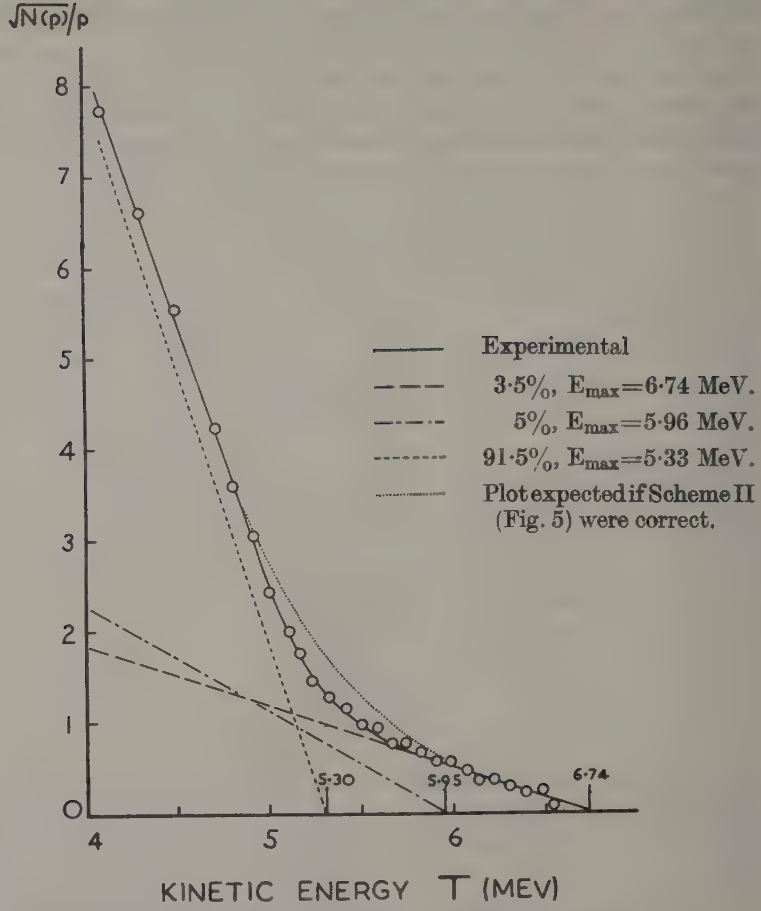
F^{20} decay schemes.

Scheme I., though possibly incomplete, is the only firmly established one. The energies given represent the best present estimates.

ACKNOWLEDGMENTS.

The author wishes to thank Dr. J. V. Jelley for his co-operation in the early experiments, Mr. E. S. Shire for the design of the original spectrometer and Mr. W. Birtwhistle for operation of the high tension equipment. He is indebted to the Department of Scientific and Industrial Research for a grant during the period of this work.

Fig. 6.



Constructed Kurie plots for the F^{20} beta ray spectrum.

References.

Feenberg, E., and Hammack, K. C., 1949, *Phys. Rev.*, **75,** 1877.

Hutchinson, G. W., 1948, *Nature, Lond.*, **162,** 610.

Konopinski, E. J., 1943, *Rev. Mod. Phys.*, **15,** 209.

Lind, D. A., Brown, J. R., and DuMond, J. W. M., 1949, *Phys. Rev.*, **76,** 1838.

Littauer, R. M., 1950, *Rev. Sci. Instrum.*, **21,** [8], 750.

Rosenfeld, L., 1948, *Nuclear Forces* (Elsivier : North Holland Publishing Co.).

Snowdon, S. C., 1950, *Phys. Rev.*, **78,** 299.

Stephens, W. E., 1934, *Phys. Rev.*, **45,** 514.

Tollestrup, A., Fowler, W. A., and Lauritsen, C. C., 1950, *Phys. Rev.*, **78,** 372.

Walker, R. L., and McDaniel, B. D., 1948, *Phys. Rev.*, **74,** 315.

CIX. *The Emission of Short Range Alpha Particles from Light Elements under Proton Bombardment.—IV. The Reaction* $^{19}F(p\alpha)^{16}O$ *.

By Joan M. Freeman,
Cavendish Laboratory, Cambridge †.

[Received August 23, 1950.]

Summary.

The short range alpha particles emitted in the proton bombardment of fluorine have been examined with a magnetic analyser for proton energies from 300 to 950 keV. In the $^{19}F(p\alpha, \gamma)^{16}O$ reaction three alpha particle groups have been measured, the average values of the energy releases being $Q_1=1{\cdot}979\pm0{\cdot}009$, $Q_2=1{\cdot}200\pm0{\cdot}010$ and $Q_3=0{\cdot}996\pm0{\cdot}010$ MeV. ; the resonances for all three groups appear to be the same. The relative intensities of the groups at six resonances have been calculated. Angular distribution measurements indicate that there are no large departures from isotropy in the cases examined. Energy measurements of the alpha particles associated with pair emission in the $^{19}F(p\alpha, \pi)^{16}O$ reaction give a value for the energy release of $2{\cdot}056\pm0{\cdot}012$ MeV.

§1. Introduction.

It is well known that the proton bombardment of fluorine can lead to the emission of short range alpha particles (subsequently designated α_1-particles) followed by gamma radiation of about 6 MeV. energy, according to the scheme :

$$\left.\begin{array}{c}^{19}F+{}^{1}H\rightarrow{}^{20}Ne^*\rightarrow{}^{16}O^*+{}^{4}He+Q_1,\\ ^{16}O^*\rightarrow{}^{16}O+\gamma_1.\end{array}\right\} \quad . \quad . \quad . \quad (1)$$

Accurate measurements of the energy of the gamma rays by Walker and McDaniel (1948) have shown that in addition to radiation of 6·13 MeV. energy a component of 6·98 MeV. radiation is also present, being directly associated with the lower energy radiation, although the ratio of the intensities of the two components varies from one resonance to another. It seemed probable that the 7 MeV. radiation was produced in a competing reaction similar to (1) in which a higher excited state of oxygen-16 was involved :

$$\left.\begin{array}{c}^{19}F+{}^{1}H\rightarrow{}^{20}Ne^*\rightarrow{}^{16}O^{**}+{}^{4}He+Q_2,\\ ^{16}O^{**}\rightarrow{}^{16}O+\gamma_2.\end{array}\right\} \quad . \quad . \quad . \quad . \quad (2)$$

---

† Communicated by Dr. W. E. Burcham.

and this was proved by observations of alpha particles of shorter range than those designated α_1 above (Burcham and Freeman 1949 a). The excitation function for these shorter range particles showed resonances in the region of proton bombarding energies from 300 to 950 keV. coinciding with resonances for the production of the α_1-particles, the intensity ratios varying considerably from one resonance to another. It was also found that the energy release Q_2 for the shorter range alpha particles was greater at the 873 keV. resonance than at the other resonances investigated. This evidence suggested the existence in the oxygen nucleus of excited levels at 7·15 and 6·94 MeV. above the ground state as well as the 6·13 MeV. level, and the pair-emitting level at about 6·00 MeV.

By improvement in the resolution of the magnetic analyser used for these investigations it has now been possible to measure the energies and intensities of the alpha particle groups more accurately and to show that at most of the resonances the shorter range alpha particles consist of two groups (designated α_2 and α_3). At the 873 keV. resonance the α_2-group, emitted in accordance with reaction (2), is the more intense, and at the other resonances studied the lower energy α_3-group, coming from

$$\left.\begin{array}{c} {}^{19}\mathrm{F}+{}^{1}\mathrm{H}\rightarrow{}^{20}\mathrm{Ne}^{*}\rightarrow{}^{16}\mathrm{O}^{***}+{}^{4}\mathrm{He}+Q_3, \\ {}^{16}\mathrm{O}^{***}\rightarrow{}^{16}\mathrm{O}+\gamma_3. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (3)$$

predominates. These groups have also been studied by Chao, Tollestrup, Fowler and Lauritsen (1950), whose results are in substantial agreement with those presented in the present paper in the cases where they overlap†. The energy of the gamma radiation (γ_2 and γ_3) associated with the α_2- and α_3-particle emission was measured carefully by Rasmussen, Hornyak, Lauritsen and Lauritsen (1950), and although the two lines were not resolved the predominant energy was shown to be less (6·99±·05 MeV.) at the 873 keV. resonance than at 939 keV. where an energy of 7·09±0·06 MeV. was found.

In the work described in subsequent sections the energies of the α_1-, α_2- and α_3-particles have been measured at a number of resonances and the corresponding Q-values deduced ; excitation functions have been obtained and the relative intensities of the groups at most of the resonances from 340 to 939 keV. have been determined. Some observations on the angular distributions of the alpha particles are also presented.

It is known that the proton bombardment of fluorine can also lead to resonance reactions in which either a long range alpha particle is emitted or a transition may occur to the lowest excited state of ^{16}O with the emission of a short range alpha particle (designated α_π) followed by an electron–position pair (see for example Rassmussen *et al.* 1950, Chao *et al.* 1950, and references included therein)

---

† We are grateful to Professor Lauritsen for communicating these results to us prior to publication.

$$\left.\begin{array}{l} {}^{19}\mathrm{F}+{}^{1}\mathrm{H}\rightarrow{}^{20}\mathrm{Ne}^{*(\pi)}+{}^{16}\mathrm{O}^{*(\pi)}+{}^{4}\mathrm{He}+\mathrm{Q}_\pi, \\ {}^{16}\mathrm{O}^{*(\pi)}\rightarrow{}^{16}\mathrm{O}+e^{+}+e^{-}. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (4)$$

Some measurements on the α_π-particles are described in section 3(*e*).

§ 2. Experimental Method.

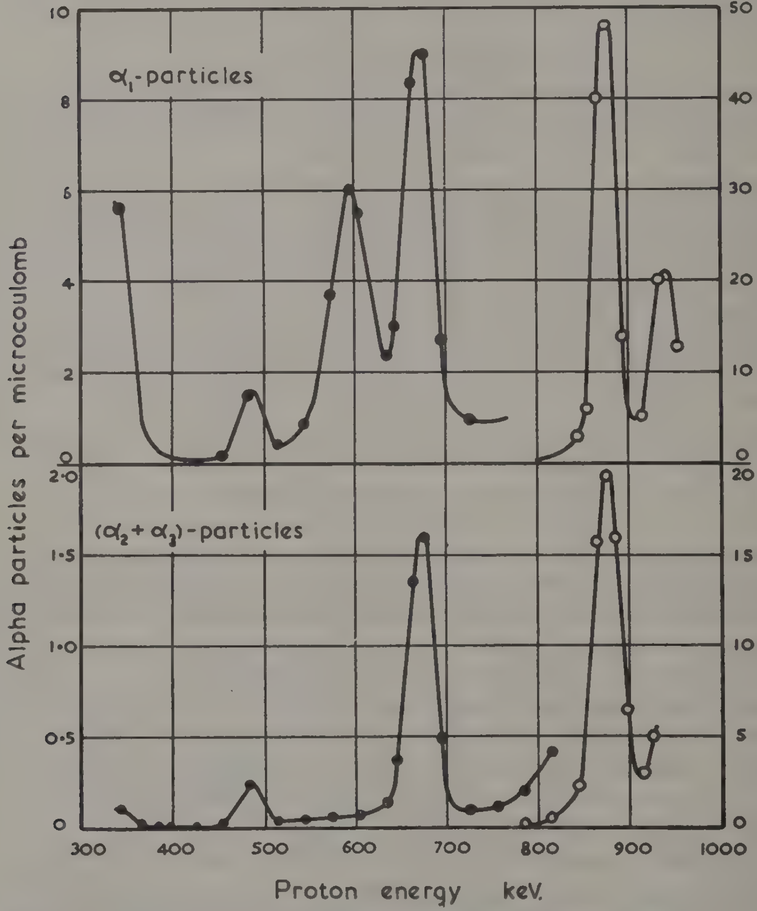
The 90° magnetic analyser used for the study of the short range alpha particle groups has already been described (Burcham and Freeman 1949 b, 1950 a, 1950 b, subsequently referred to as I., II. and III.). In much of the work the magnet was used on its side so that observations could be made in a horizontal plane ; the proton beam which reached the target could be restricted to a diameter of 5 mm. and the zinc sulphide scintillation screen used with an EMI type 4588 photo-multiplier for detecting the analysed particles was then made 5 mm. wide in the focusing plane. Under these conditions the effective resolution of the system (ΔE/E) was 4 per cent in energy. The analyser was calibrated by observations on the peaks of proton and deuteron groups of known energy scattered from a thin gold film, and of deuteron and alpha particle groups from the proton bombardment of beryllium (*cf.* I. ; the Q values used for the ${}^{9}\mathrm{Be}(pd){}^{8}\mathrm{Be}$ and ${}^{9}\mathrm{Be}(p\alpha){}^{6}\mathrm{Li}$ reactions were 0·558 and 2·121 MeV. respectively (Tollestrup, Fowler and Lauritsen 1949)).

The voltage scale of the 1 MV. accelerating equipment set used for the experiments was calibrated by reference to the gamma ray excitation function for the fluorine reaction (1), observations being made with a thick sodium fluoride target, using a Geiger counter. The relative positions of the gamma ray resonances have been determined by the work of Bennett *et al.* (1946) and Bonner and Evans (1948), and the absolute values were estimated by correcting their value of 862 keV. for the most prominent resonance in the region explored to the figure of 873·5 keV. found by Herb, Snowdon and Sala (1949), and multiplying their figures for the other resonances by the same factor.

The fluorine targets used for the alpha particle work were prepared by evaporating thin layers of calcium fluoride on copper buttons in vacuo ; the surface densities of the layers were estimated by weighing pieces of mica placed close to the buttons during the evaporation process. Target thickness effects were calculated on the assumption that the stopping power of calcium fluoride for protons of 500 keV. energy was 260 keV./mgm./cm.^{2} and that for alpha particles of 1 and 2 MeV. energy it was 1600 and 1100 keV./mgm./cm.^{2} respectively. In most cases targets of thickness 0·02 mgm./cm.^{2} or less were used.

In the experiments on the alpha particles associated with gamma ray emission the yield was monitored by means of a Geiger counter, mounted vertically below the target ; simultaneous measurements of alpha particle and gamma-ray intensities were made, the background of the Geiger counter being first determined by observations with a blank target. Counts were usually made for a standard number of microcoulombs collected by the target, as indicated by a current integrator.

Fig. 1.



Excitation functions for the α_1-particles and the unresolved $(\alpha_2+\alpha_3)$-particles from the $^{19}F(p\alpha, \gamma)^{16}O$ reaction. Left-hand ordinates refer to the region below 800 keV. proton bombarding energy, and right-hand ordinates to that above 800 keV.

§ 3. Experimental Results.

(a) *Alpha particles associated with gamma radiation.*

In the first experiments (Burcham and Freeman 1949 a), in which an analyser resolution of 10 per cent in energy was used, the lower energy (α_2- and α_3-) particles appeared as a single group. The excitation functions for these and for the α_1-particles were observed at the same time by measuring the intensity at the peak of each group as a function of proton bombarding energy from 330 to 950 keV. Observations were made at an angle of 83° with the proton beam. The curves are reproduced in fig. 1, in which the proton energy scale has been adjusted to place the most prominent resonance in the α_1-particle curve at the value 873·5 keV. The widths of the resonances are about 30 keV., which is equal to the estimated spread in energy of the bombarding proton beam, except in the case of the 598 keV. resonance in the α_1-particle curve which is slightly broader, as expected from the observations of Burcham and Devons (1939), and Bonner and Evans (1948). All the resonances observed for the lower energy alpha particles appear to coincide exactly with the positions of α_1-particle resonances and it may be concluded that the same excited levels of the compound $^{20}\text{Ne}^*$ nucleus are involved in each case, although the branching ratios vary with proton energy. There is no resonance in the curve for α_2- and α_3-particles corresponding to the 598 keV. resonance for α_1-particles.

For more accurate energy measurements an analyser resolution of 4 per cent in energy was used, and the alpha particles were observed at an angle of 135° with the proton beam. Fig. 2(*a–f*) shows the alpha particle spectra obtained with targets of thickness $\sim$0·02 mgm./cm.^{2} at the 939, 873, 669, 598, 486 and 340 keV. resonances. Excepting the case of the 598 keV. resonance, where only the α_1-particles appear, there is evidence for both α_2- and α_3-particle groups at each resonance, although for the lower resonances the position and intensity of the weakest (α_2) group cannot be fixed very accurately. During the course of the experiments in which these spectra were obtained a film of carbon contamination formed on the target surface. The development of this contamination was traced by observations from time to time on the position of the peak of the α_1-particle group at the 873 keV. resonance and the corrections to be applied to the observed alpha particle energies were so obtained. The spectra shown in fig. 2 have been corrected in this way. Targets were frequently changed in order to avoid the necessity for large corrections for the carbon deposit. The effective target thicknesses were estimated to be equivalent to about 22 keV. for 2 MeV. alpha particles or 32 keV. for 1 MeV. alpha particles, so that corrections for target thickness were small. This was verified by observations of the extrapolations of the high energy sides of the spectra as described in II. and III.

From the corrected energies of the three sets of alpha particle groups the corresponding Q-values for reactions (1), (2) and (3) were obtained, and these are given in Table I. The probable erros are calculated from the estimated uncertainties in the energy calibration of the analyser and in the target thickness and carbon contamination corrections. The mean values for Q_1, Q_2 and Q_3 were estimated by weighting the figures for each resonance according to the inverse squares of the probable errors

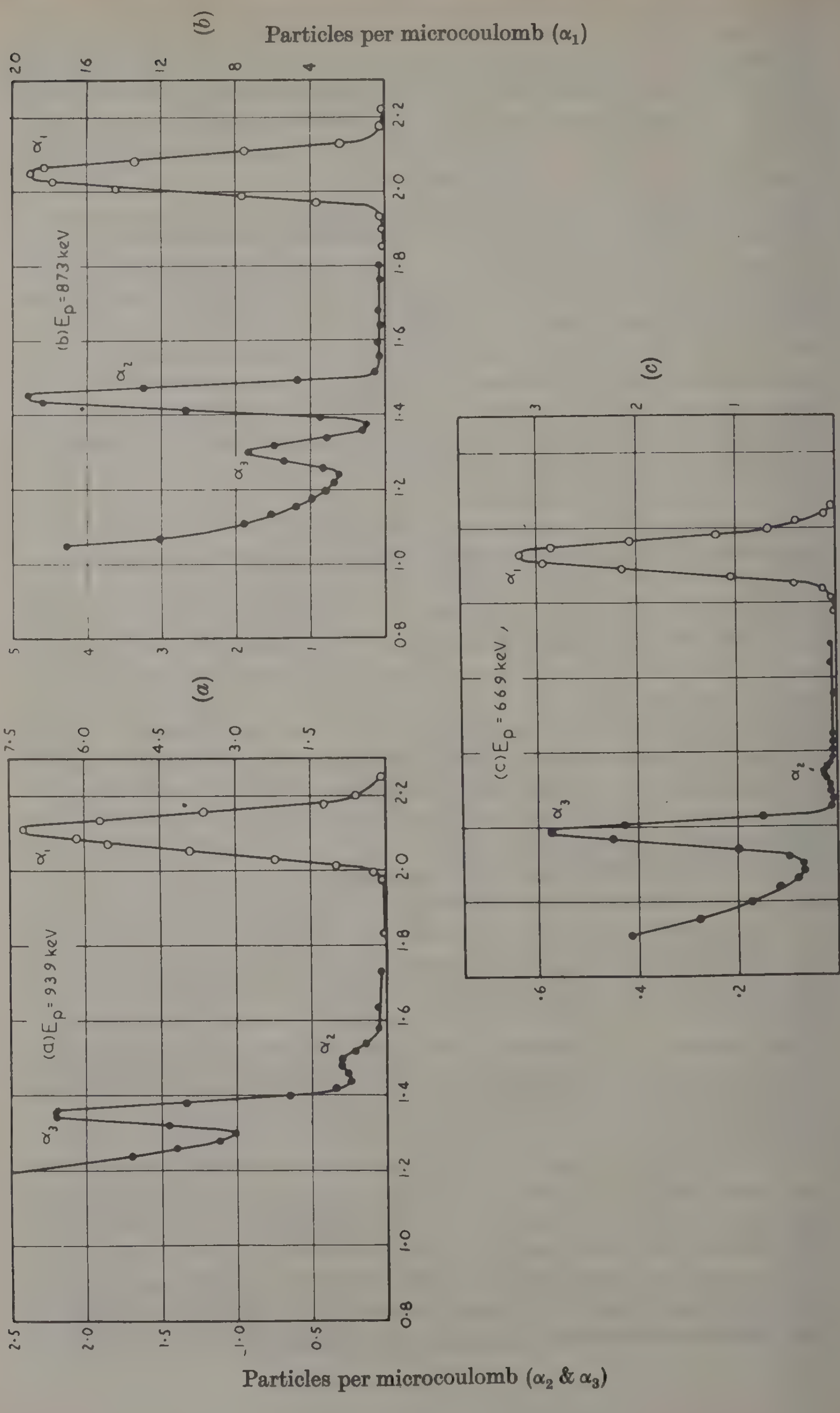
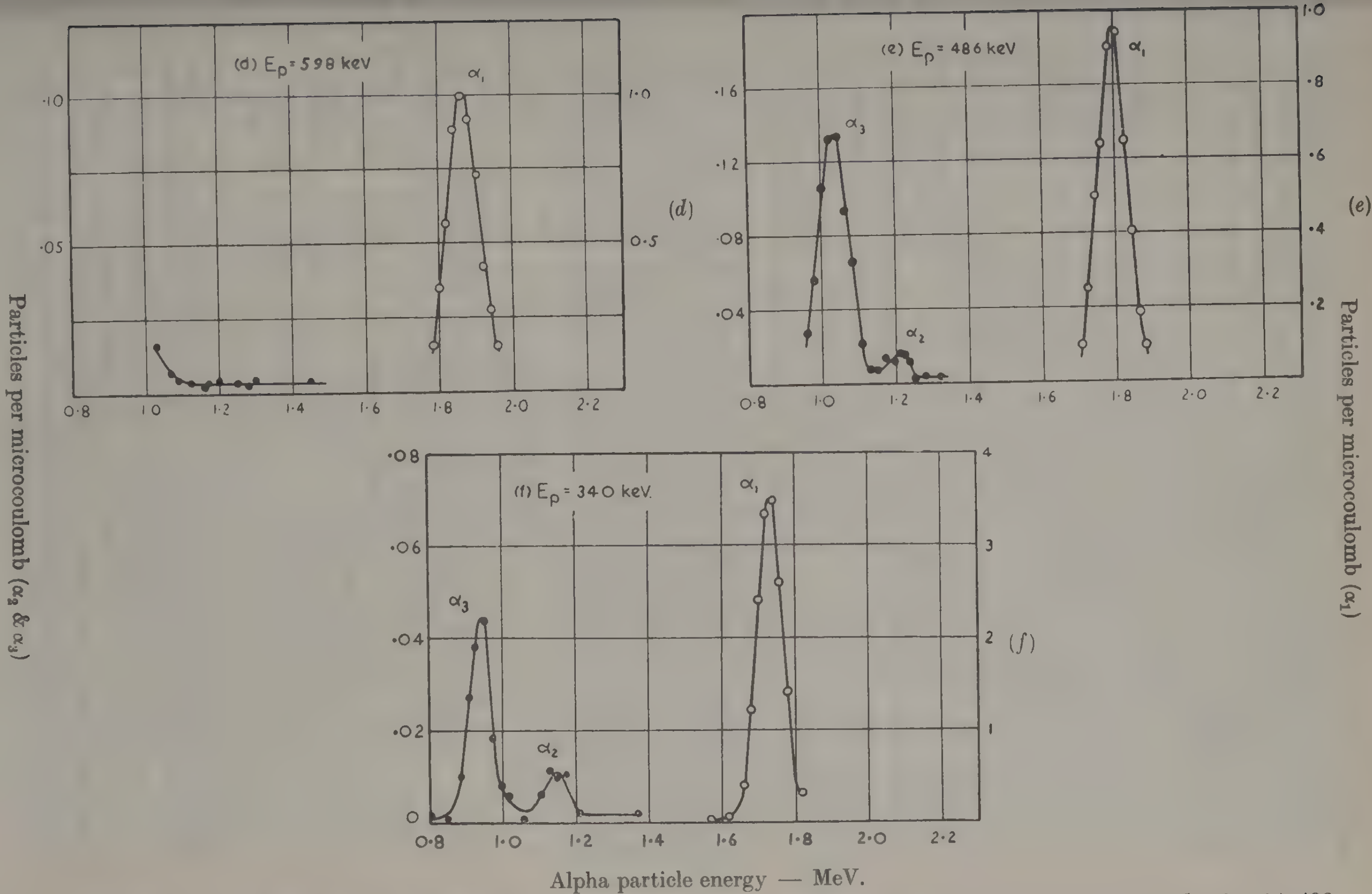

(b)
Particles per microcoulomb (α_1)
20
16
12
8
4
5
4
3
2
1
(b) E_p = 873 keV
α_1
α_2
α_3
0·8
1·0
1·2
1·4
1·6
1·8
2·0
2·2
(a)
7·5
6·0
4·5
3·0
1·5
2·5
2·0
1·5
1·0
0·5
(a) E_p = 939 keV
α_1
α_2
α_3
0·8
1·0
1·2
1·4
1·6
1·8
2·0
2·2
(c)
3
2
1
·6
·4
·2
(c) E_p = 669 keV
α_1
α_2
α_3
0·8
1·0
1·2
1·4
1·6
1·8
2·0
2·2
Particles per microcoulomb (α_2 & α_3)

Fig. 2 (*a–c*).



Alpha particle spectra from the $^{19}F(p\alpha, \gamma)^{16}O$ reaction at proton energies of (*a*) 939, (*b*) 873, (*c*) 669, (*d*) 598, (*e*) 486 and (*f*) 340 keV. Right-hand ordinates give yields of α_1-particles per microcoulomb of protons and left-hand ordinates give the α_2- and α_3-particle yields in the same units. Abscissæ give alpha particle energies in MeV.

quoted. The values obtained ($Q_1=1{\cdot}979\pm0{\cdot}009$, $Q_2=1{\cdot}200\pm0{\cdot}010$, $Q_3=0{\cdot}996\pm0{\cdot}010$ MeV.) are in very good agreement with the figures deduced by Chao *et al.* (1950) from measurements at the 873 and 1353 keV. resonances ($Q_1=1{\cdot}977\pm0{\cdot}008$, $Q_2=1{\cdot}204\pm0{\cdot}008$ and $Q_3=1{\cdot}002\pm0{\cdot}008$ MeV.) and with the value $Q_1=1{\cdot}969$ MeV. given by Strait, Van Patter and Buechner (1950).

TABLE I.

E_p (keV.)	Q_1 (keV.)	Probable error (keV.)	Q_2 (keV.)	Probable error (keV.)	Q_3 (keV.)	Probable error (keV.)
939	1991	20	1199	20	999	20
873	1980	14	1195	14	996	14
669	1984	20	1219	25	1017	20
598	1959	25	—	—	—	—
486	1962	35	1191	35	973	35
340	1979	20	1207	25	971	25
Weighted mean	1979	9	1200	10	996	10

The relative yields at 135° of α_1-, α_2- and α_3-particles, at each of the resonances studied, were obtained from the spectra shown in fig. 2 by calculating for each group the integral $\int N(E)dE/E$, where $N(E)$ is the number of alpha particles counted at a given setting of the field current of the magnetic analyser, corresponding to the energy E. A correction for the number of singly-charged alpha particles expected to be associated with each of the observed doubly-charged groups was made using the data given by Briggs (1927) and Knipp and Teller (1941). The values thus obtained for the ratios α_1/α_2, α_1/α_3 and $\alpha_1/(\alpha_2+\alpha_3)$ at each resonance are set out in Table II. In the cases where the ratios are large the figures must be regarded as rough approximations only. An allowance for scattered protons had to be made in estimating the yield of the α_3-group at the 939 keV. resonance. In the region where the

TABLE II.

E_p (keV.)	$\frac{\alpha_1}{\alpha_2}$	$\frac{\alpha_1}{\alpha_3}$	$\frac{\alpha_1}{(\alpha_2+\alpha_3)}$
939	26	3·3	2·9
873	3·8	8·6	2·6
669	250	4·7	4·6
598	>400	>400	200
486	65	4·0	3·8
340	250	50	40

results in Table II. overlap the observations of Chao *et al.* (1950), for an angle of 138°, a comparison shows general but not complete agreement.

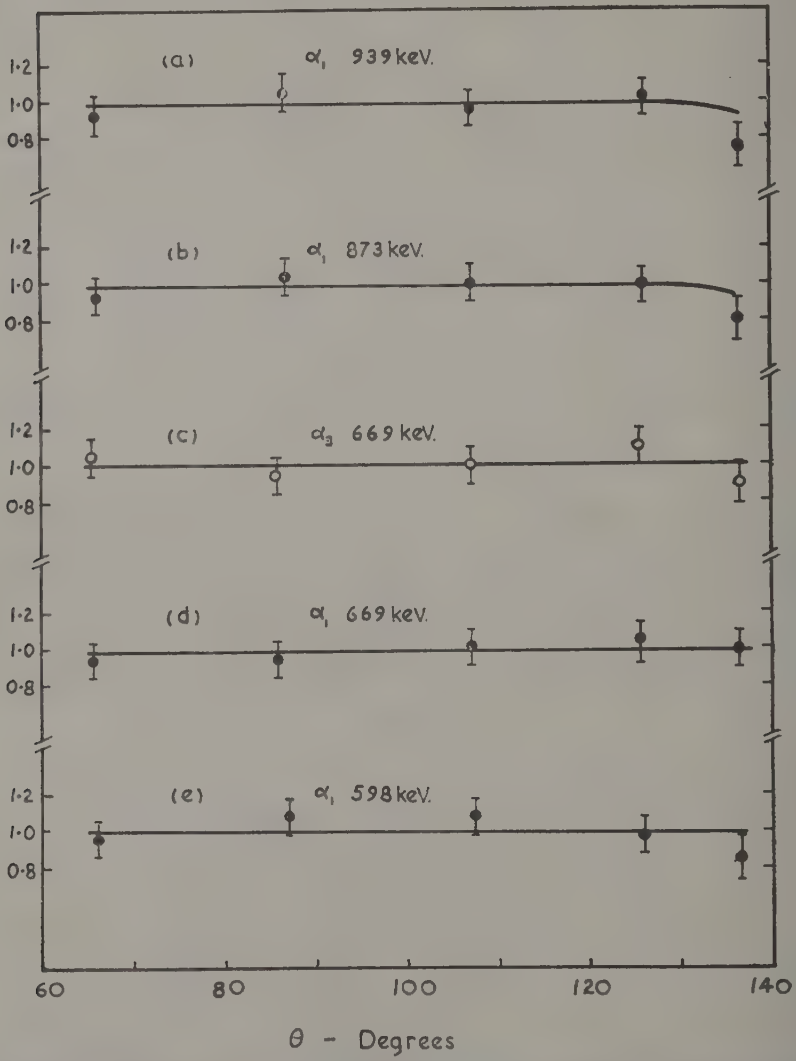
(b) *Angular distributions.*

The angular distributions of the alpha particles associated with gamma-ray emission have been examined at several resonances, over a limited range of angles. The experimental arrangement, which has already been described (1, § 2 (*d*)), permitted observations to be made at angles between 60° and 135°. Thin targets were used and the variation in the effective target thickness as the target was rotated for each angular position was taken into account by the use of the gamma ray monitoring counter mounted vertically below the target. Observations were made of the ratio of the alpha particle counting rate at the peak of the spectrum under examination to the simultaneous gamma ray counting rate (corrected for background). The distributions of α_1-particles from the 939, 873, 669, 598 and 340 keV. resonances were observed, and for the energy of 669 keV. the α_3-particle distribution was also studied. When corrected to the centre of mass system the readings at all resonances showed deviations from isotropy slightly outside the statistical accuracy, which was about 10 per cent. Now the distribution of α_1-particles emitted at the 340 keV. resonance has been examined by Van Allen and Smith (1941), and shown to be spherically symmetrical to within 2 per cent. It was therefore assumed that the distribution obtained in the present experiment at this resonance actually corresponded to an isotropic distribution and that apparent variations in yield with angle were due to systematic errors which would affect the observations at all resonances in the same way. The 340 keV. results were therefore used to correct the observations made at the other resonances; for each case the yield observed at a given angle θ was divided by the yield of α_1-particles obtained from the 340 keV. resonance at the same angle θ. The distributions derived in this way are shown in fig. 3; suitable normalizing factors have been applied to make the average value of the ratio $I_\theta(E)/I_\theta(340$ keV.) approximately unity in each case. It can be seen that, within the experimental accuracy (about 10 per cent) and over the limited range of angles which it was possible to cover, no significant departures from isotropy are apparent in any of the cases examined. This result is in agreement with conclusions of Chao *et al.* (1950) derived on a different basis: they measured, for a number of resonances, the yields of α-$_1$-, α_2- and α_3-particles from a given target at a particular angle (138°); the total yields over 4π steradians were then calculated on the assumption of isotropic distributions of all the particles, and these figures were found to agree to within 10 per cent with total yields of gamma radiation measured simultaneously.

(c) *Alpha particles associated with pair emission.*

The α_π-particles emitted according to the reaction (4), § 1, involving the lowest excited state of ^{16}O, were studied in the neighbourhood of the resonance at 843 keV., which is known from measurements of the electron–positron pairs (Bennett *et al.* 1946). The alpha particle spectrum obtained at an angle of observation of 135° and a bombarding energy

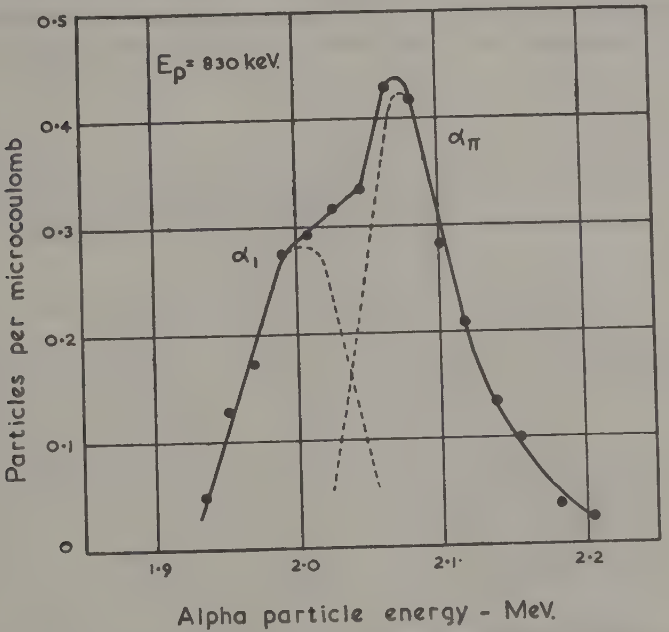
Fig. 3.



Angular distributions of the alpha particles from the $^{19}F(p\alpha, \gamma)^{16}O$ reaction: α_1-patricles at proton energies of (*a*) 939, (*b*) 873, (*d*) 669, (*e*) 598 keV.: curve (*c*) is for α_3-particles at 669 keV.

of 830 keV. is shown in fig. 4. The target was 0·02 mgm./cm.^{2} thick and the analyser resolution 4 per cent. The presence of two groups is evident although complete resolution was not obtained. The lower energy group shown in the figure has an energy which would correspond to α_1-particles from the " tail " of the 873 keV. resonance studied above. The higher energy group (2·08 MeV.) is attributed to the α_π-particles preceding pair emission according to (4). A similar spectrum was obtained at proton energy of 820 keV., and from the two measurements a mean Q-value for reaction (4) was deduced as $Q_\pi = 2{\cdot}056 \pm 0{\cdot}012$ MeV. This result is in agreement with the value of $2{\cdot}061 \pm 0{\cdot}010$ MeV. found by Chao *et al.* (1950).

Fig. 4.

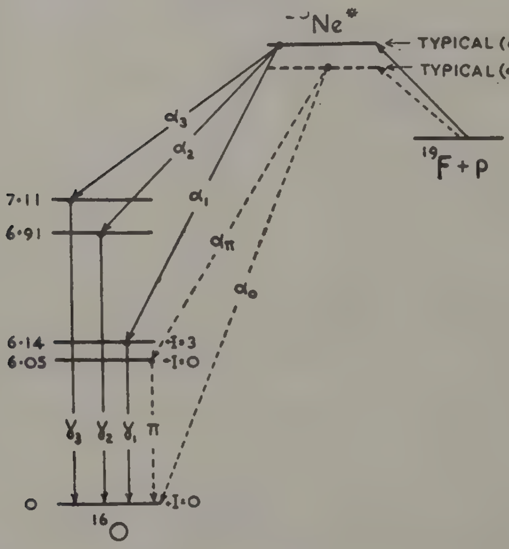


Alpha particles from the reactions $^{19}F(p\alpha, \gamma)^{16}O$ and $^{19}F(p\alpha, \pi)^{16}O$ at a proton bombarding energy of 830 keV.

§ 4. Discussion.

From the present state of knowledge concerning reactions produced by the proton bombardment of fluorine one can draw up a generalized disintegration scheme as shown in fig. 5. The excited states formed in ^{20}Ne appear to fall into two classes : (i) levels from which transitions may occur either to the ground state of ^{16}O with long range alpha particle (α_0) emission or to the lowest excited state of ^{16}O with short range alpha particle (α_π) emission followed by the emission of an electron-positron pair (π) ; (ii) levels from which transitions may occur to any one of the next three excited levels of ^{16}O with emission of short range alpha particles (α_1, α_2 or α_3) followed by gamma radiation (γ_1, γ_2 or γ_3). In the latter case the relative probabilities of the transitions to the three ^{16}O levels vary considerably from one resonance to another as is shown by the intensity ratios listed in Table II. An attempt was made to determine the way in which these relative probabilities might be governed by the penetrability factors for the emitted alpha particles, whose energies are well below the barrier of the residual ^{16}O nucleus. The penetrability factors were calculated from the formulæ given by Bethe (1937), assuming a constant value for the radius of the ^{16}O nucleus derived from the data of Amaldi and Cacciapuoti (1947). The factors appropriate to alpha particles emitted with angular momenta of all values up to $l=4$ were calculated and an attempt was made to select consistent sets of the ratios of these factors which would fit the observed α_1/α_2 and α_1/α_3 ratios at the various resonances, and give some information about the angular momenta of the emitted particles and the spins of the $^{16}O^*$ levels. No such self-consistent sets could however be found. One is tempted to conclude, in a general way, that in the cases where the ratios α_1/α_2 or α_1/α_3 are fairly small the angular momenta of the α_1-particles are greater than those of the lower energy particles, so that the emission probabilities of the latter are raised to comparable values. On the other hand, in the case of the 598 keV. resonance the high relative probability of α_1-particle emission could be accounted for if these particles were assumed to be emitted with small or zero angular momentum while the lower energy

Fig. 5.



Disintegration modes in the proton bombardment of fluorine. The values (in MeV.) assigned to the ^{16}O levels are derived form the value $Q_0=8{\cdot}11$ MeV given by Chao *et al.* (1950).

particles encountered an appreciable angular momentum barrier. Such an assumption would also account for the greater width of the 598 keV. resonance compared with the other resonances (*cf.* Devons and Hine 1949).

The recent angular correlation measurements made in this laboratory for the α_1-particles and γ_1-radiation from the 340 keV. resonance (Barnes, French and Devons 1950) indicate a useful method of obtaining data on the angular momenta of the α-particles from the various $^{20}Ne^*$ levels, and the spins of these and the $^{16}O^*$ levels. When such data have been obtained, further deductions may then be possible from the observed intensity ratios discussed above. It may appear, for example, that the radius of the $^{16}O^*$ nucleus is not the same for the three levels of excitation involved.

ACKNOWLEDGMENTS.

I should like to express my thanks to Dr. W. E. Burcham for much advice and assistance in this work. I also wish to thank Messrs. W. Birtwhistle and D. D. Stewart for much technical assistance

REFERENCES.

AMALDI, E., and CACCIAPUOTI, B. N., 1947, *Phys. Rev.*, **71,** 739.
BARNES, C. A., FRENCH, A. P., and DEVONS, S., 1950, *Nature, Lond.*, **166,** 145.
BENNETT, W. E., BONNER, T. W., MANDEVILLE, C. E., and WATT, B. E., 1946, *Phys. Rev.*, **70,** 882.
BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9,** 166, 178.
BONNER, T. W., and EVANS, J. E., 1948, *Phys. Rev.*, **73,** 666.
BRIGGS, G. H., 1927, *Proc. Roy. Soc.* A, **114,** 341.
BURCHAM, W. E., and DEVONS, S., 1939, *Proc. Roy. Soc.* A, **173,** 555.
BURCHAM, W. E., and FREEMAN, J. M., 1949 a, *Phys. Rev.*, **75,** 1756; 1949 b, *Phil. Mag.*, **40,** 807 ; 1950 a, *Ibid.*, **41,** 337 ; 1950 b, *Ibid.*, **41,** 921.
CHAO, C. Y., TOLLESTRUP, A. V., FOWLER, W. A., and LAURITSEN, C. C., 1950, *Phys. Rev.*, **79,** 108.
DEVONS, S., and HINE, M. G. N., 1949, *Proc. Roy. Soc.* A, **199,** 56.
HERB, R. G., SNOWDON, S. C., and SALA, O., 1949, *Phys. Rev.*, **75,** 246.
KNIPP, J., and TELLER, E., 1941, *Phys. Rev.*, **59,** 659.
RASMUSSEN, V. K., HORNYAK, W. F., LAURITSEN, C. C., and LAURITSEN, T., 1950, *Phys. Rev.*, **77,** 617.
STRAIT, E. N., VAN PATTER, D. M., and BUECHNER, W. W., 1950, *Bull. Amer. Phys. Soc.*, **2255,** No. 1, 33.
TOLLESTRUP, A. V., FOWLER, W. A., and LAURITSEN, C. C., 1949, *Phys. Rev.*, **76,** 428.
VAN ALLAN, J. A., and SMITH, N. M., 1941, *Phys. Rev.*, **59,** 501.
WALKER, R. L., and MCDANIEL, B. D., 1948, *Phys. Rev.*, **74,** 315.

CX. *An Interferometric Study of some Optical Properties of Evaporated Silver Films.*

By R. C. **Faust**, Ph.D.*,
Royal Holloway College †.

[Received August 23, 1950.]

[Plates XXXVI. & XXXVII.]

Abstract.

Thermally evaporated silver films have been studied in the reflectivity range of 4 to 95 per cent. The air-silver and glass–silver reflection and transmission coefficients, the corresponding reflection phase changes and the apparent film thickness have been measured and compared with those of sputtered silver films. For very thin films, in a region where the light absorption is high, the glass-silver reflection coefficient passes through a minimum, and in the yellow and the green, but not in the blue, the glass–silver reflection phase change suddenly alters from an advance to a retardation. This transition in the phase change is intimately associated with the reflected fringe asymmetry.

§ 1. Introduction.

The presence of metallic coatings on the plates of an interferometer modifies the fringe patterns, the reflected system being affected differently from the transmitted one. In fig. 1 are seen three isotropic dielectric media 1, 2, 3, separated by the two parallel planes XX′ and YY′ which may or may not be coated with metallic films. The central medium has a refractive index μ and a geometrical thickness t. Monochromatic radiation which is incident upon XX′ enters medium 2 at an angle θ with the normal, and gives rise to a transmitted series of parallel wave trains $a, b, c, \ldots$, and a reflected series A, B, C, ... The optical properties of the boundary XX′ for light incident from medium 1 are an amplitude reflection coefficient r_1, an amplitude transmission coefficient t_1, a reflection phase change α_1 and a transmission phase change β_1 ; if the light is incident upon XX′ from medium 2 the same symbols are used with the suffix 2. If the light falls upon YY′ from medium 2 the reflection and transmission coefficients are ρ and τ, whilst the reflection phase change is γ. These optical values, which apply to light of vacuum wavelength λ incident upon the interferometer at one particular angle, are in general dependent upon the state of polarization.

---

* Communicated by Professor S. Tolansky.

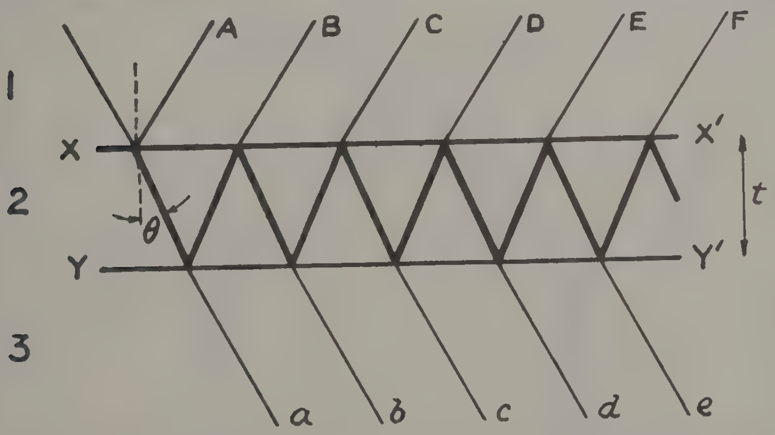
† Now at the British Rayon Research Association, Urmston.

The resultant intensity of the transmitted light is a periodic and symmetrical function of the phase quantity

$$\delta = 4\pi/\lambda \, . \, \mu t \cos\theta + \alpha_2 + \gamma \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

the greatest and least intensities occurring at $\delta = 2n\pi$ and $(2n+1)\pi$ respectively. It has, however, been shown by Hamy (1906) that the reflected intensity is a periodic but generally asymmetrical function of δ. A given reflection minimum is therefore not necessarily equidistant (in terms of δ) from the two neighbouring reflection maxima, nor does the condition for a reflection minimum generally correspond to that leading to a transmission maximum. It is to be noted, however, that whenever XX′ is an uncoated dielectric interface the reflected fringe pattern is symmetrical, the reflected intensity minima coinciding with the transmitted maxima.

Fig. 1.

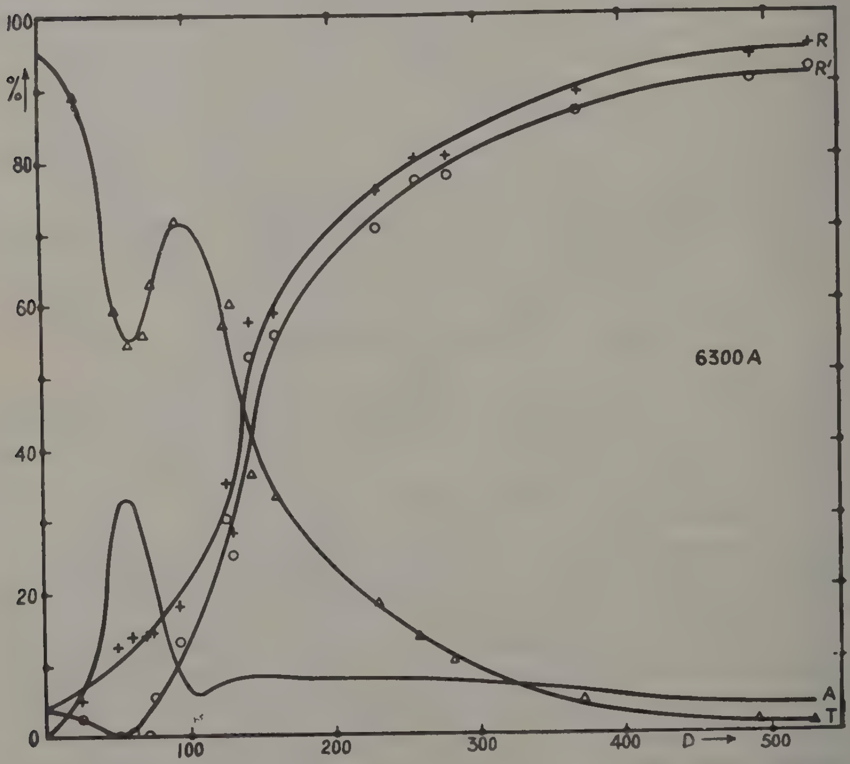


Hamy confirmed his theoretical predictions by observing the reflected fringe system produced by interferometers whose plates had been coated with chemically deposited silver films. More recently silver films evaporated on to glass have been similarly studied by Holden (1949), who, having compared his findings with the optical properties of sputtered films as determined by Rouard (1937), suggested that certain features of the reflection fringe asymmetry are closely associated with the glass silver reflection phase change. In order to test this idea the properties of evaporated silver films have been examined at normal incidence and compared with those of sputtered films. The apparent film thickness, the reflection and transmission coefficients and the reflection phase changes at both the air–silver and glass–silver interfaces were measured and the results related to the observed reflection fringe asymmetry.

The silver, evaporated from molybdenum strip filaments at a pressure of about 5×10^{-5} mm. Hg, was deposited on to two glass optical flats and a thin glass coverslip. The variation in the mass deposited upon unit surface area was so slight that the three silverings could be regarded as identical. The two flats formed an air-gap interferometer, whilst the coverslip was employed for the determination of the glass–silver reflection phase change. All the measurements were made in the space of a few hours during which time no significant changes in the films' properties or appearance were detected.

The electron microscope has revealed a discontinuous structure in semi-transparent metal films, and thus one may not strictly speak of the thickness of a film. The interferometric method [Khamsavi and Donaldson (1947), Avery (1949)] which was employed yields an apparent thickness D, which is that the film would have if it were uniformly distributed and of the same density as the bulk metal. A strip of the evaporated film was gently removed with a razor blade and an opaque silver film deposited over the whole surface. the height of the well-defined step being determined from reflected Fizeau fringes. The probable error in the mean step height did not exceed ±15 A. for apparent film thicknesses in excess of 100 A. and ±25 A. for thinner films.

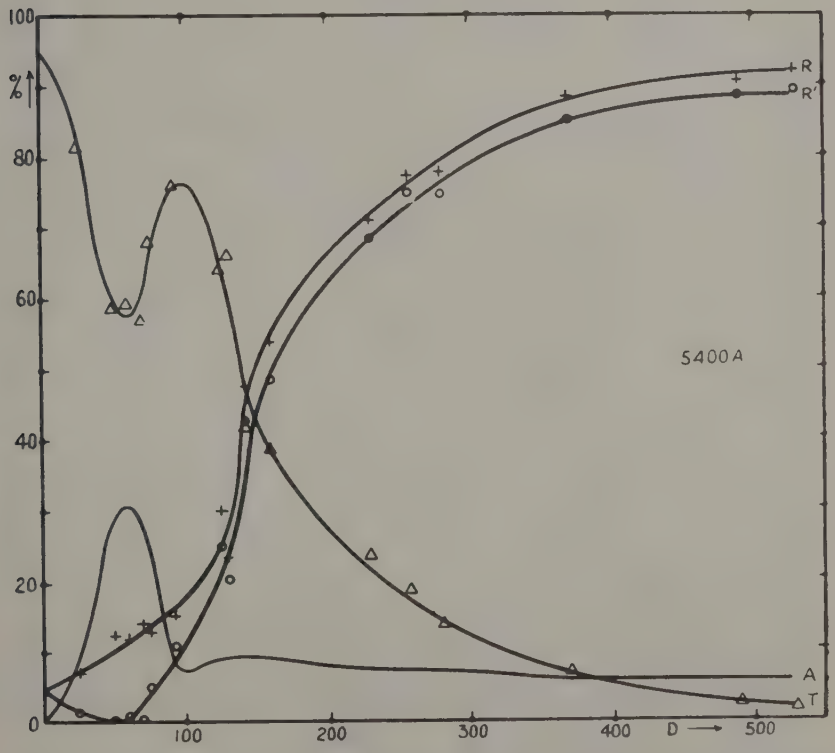
Fig. 2 (*a*).



§ 2. The Reflection and Transmission Coefficients.

An electron multiplier cell was used to compare the incident light intensity with the intensities of the beams reflected and transmitted by the silver films, the angle of incidence being only 4°. After the introduction of correction terms necessitated by the presence of the glass substitute, the desired air–silver and glass–silver reflection and transmission coefficients were calculated. Three wavelength bands were selected by filters; 6700 A. to 5900 A. with a peak at 6300 A., 5500 A. to 5200 A. with a peak at 5400 A., and 4650 A. to 4350 A. with a peak at 4500 A.

Fig. 2 (*b*).

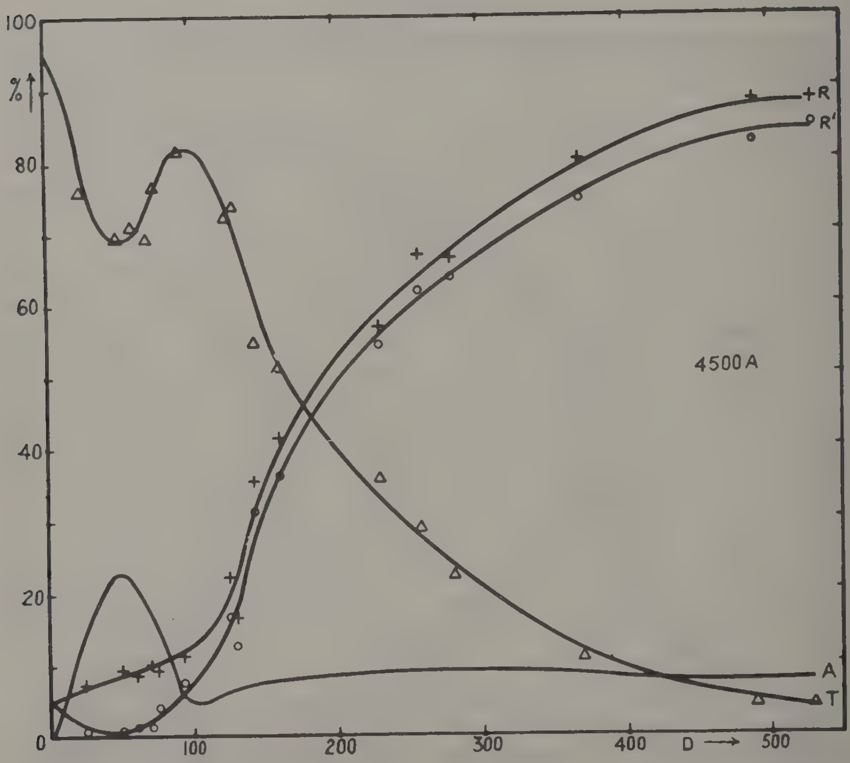


The results are presented graphically in fig. 2. As the air–silver and glass–silver transmission coefficients agreed to within the estimated error (±0·02 T), a mean value T has been taken. The error in the air-silver reflection coefficient R is about ±0·01 R for high reflectivities, but increases to ±0·04 R when the reflectivity is below 10 per cent. The inaccuracies in the glass–silver reflection coefficient R′ are greater, rising from ±0·02 R′ to as much as ±0·5 R′ at the reflectivity minimum.

The absorption for light incident from the air side A and from the glass side A′ have been calculated from the measured quantities, the presence of any light scattering being neglected. Being difference terms their errors are relatively high and small fluctuations in the calculated absorptions are not necessarily significant.

These results have been compared with those obtained by Goos (1936) and Rouard (1937) for sputtered films, and by Krautkrämer (1938), Strong and Dibble (1940), and Sennett and Scott (1950) for evaporated silver films. The following apparent general conclusions may be drawn :—

Fig. 2 (*c*).



(1) At a given wavelength the air–silver reflection coefficient R decreases with diminishing film thickness until the value appropriate to the substrate alone is reached.

(2) The wavelength variation of R depends upon the film thickness. Normally R increases with increasing wavelength, but with very thin films the opposite behaviour has been found by Goos, Sennett and Scott, and the writer, whose 25 A. thick film had the coefficients 5·0 per cent. (6500 A.), 6·9 per cent. (5400 A.) and 7·4 per cent (4500 A.).

(3) At a given wavelength the substrate-silver reflection coefficient R′ decreases with diminishing film thickness until a low minimum value is reached ; it then increases finally attaining the value appropriate to the uncoated substrate. R′ is apparently always less than R.

(4) For thick films R′ rises with increasing wavelength, but for thin films in the neighbourhood of the R′ minimum the reverse may occur, this behaviour being responsible for the beautiful reflection colours observable in thin films. The 50 A. film, for example, was blue-mauve in reflection from the glass side, R′ being greater in the blue than in the red. The 25 A. film appeared rose-pink, R′ now being weaker in the blue ; the reflection colour from the air side of this film was a brilliant torquoise because R possessed the relatively uncommon property of being strongest in the blue.

(5) The transmission measurements made by the different workers show little agreement, the major differences occurring in the neighbourhood of the substrate-silver reflectivity minimum. In this region the writer has found that the general increase in transmission with diminishing film thickness is interrupted by a marked local minimum. This was not observed by Goos or by Strong and Dibble. (Rouard and Krautkrämer did not take transmission measurements in this region.) The recent work of Sennett and Scott is very illuminating ; they found a minimum with high but not low rates of evaporation, the relevant factor being the film structure.

(6) Whether there is a minimum in the transmission curve or not, the absorption terms A and A′ have peak values for all wavelengths in the neighbourhood of the substrate-silver reflectivity minimum. The peak absorption for both A and A′ decreases with shortening wavelength, but for the purpose of clarity A′, which is everywhere greater than A, has not been plotted. According to Sennett and Scott the slower the rate of absorption, the greater the film thickness for which the peak occurs, the onset of this high absorption being closely related to the state of aggregation of the film. For evaporation times of less than a minute the maximum absorption was found for an approximate film thickness of 100 A. ; the figure obtained by the writer using high evaporation rates was 60 A. (The slight dip near the peak of the absorption curve lies within the experimental errors.)

It is thus evident that the exact mode of film preparation plays a very important role in determining the film properties, and quantitative comparisons between sets of films made under different conditions are unsatisfactory. The general qualitative behaviour described above is, however, apparently common to all sets of silver films, evaporated or sputtered.

§ 3. OBSERVATION OF THE FRINGE PATTERNS.

The theory of reflection fringe asymmetry was originally developed for a plane-parallel sheet interferometer, but both v. d. Pahlen (1912) and Brossel (1947) have shown that provided the gap thickness and the slope of a wedge interferometer are small, the fringe intensity distribution on a Feussner surface differs negligibly from that produced by a sheet interferometer. A wedge can therefore be utilized to study the fringe patterns produced by metal-coated interferometer plates, any asymmetry due to the wedge slope being insignificant. There are two zero order Feussner surfaces on which the fringes lie, one in transmission, the other in reflection, but they do not exactly coincide even when the light falls upon the interferometer at normal incidence. In order therefore that both surfaces could be focused simultaneously a well-collimated normally incident beam, an interferometer of small gap and slope, and a 2-inch (N.A. 0·17) objective were used.

Fizeau fringes and fringes of equal chromatic order (Tolansky 1945) were studied in both transmission and reflection. The observation of the fringes of equal chromatic order was carried out, using a specially constructed plane grating spectrograph (14,200 lines/inch, inverse dispersion 50 A./mm.). The spectrum was normal, wavelength differences being proportional to the linear displacements on the photographic plate.

§ 4. THE AIR-SILVER REFLECTION PHASE CHANGE.

Consider an air-gap interferometer with an unsilvered front surface XX′ and a rear surface YY′, of which one half is silvered. The reflection Fizeau fringes, the minima of which coincide with the transmission maxima, run from the glass/glass to the glass/silver region and at the junction a fringe displacement is observed (figs. 11, 13, Pl. XXXVI.). At normal incidence the positions of the reflection minima in the two regions are determined by the equation

$$n\lambda=2t+\Omega\lambda=2t'+\Omega'\lambda, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

in which the dash notation applies to the unsilvered region. This is derived from equation (1), t and t' being the two gap thicknesses at which a fringe of given order n occurs. Ω and Ω', representing the phase terms, are related by $(\Omega-\Omega')=1/2\pi\,.\,(\gamma-\gamma')$, where γ and γ' are the phase changes (in radians) at the air-silver and the air–glass surfaces respectively.

If $dq\lambda$ represents twice the silver film thickness and dp the observed fractional order shift of the two fringe systems, then one obtains

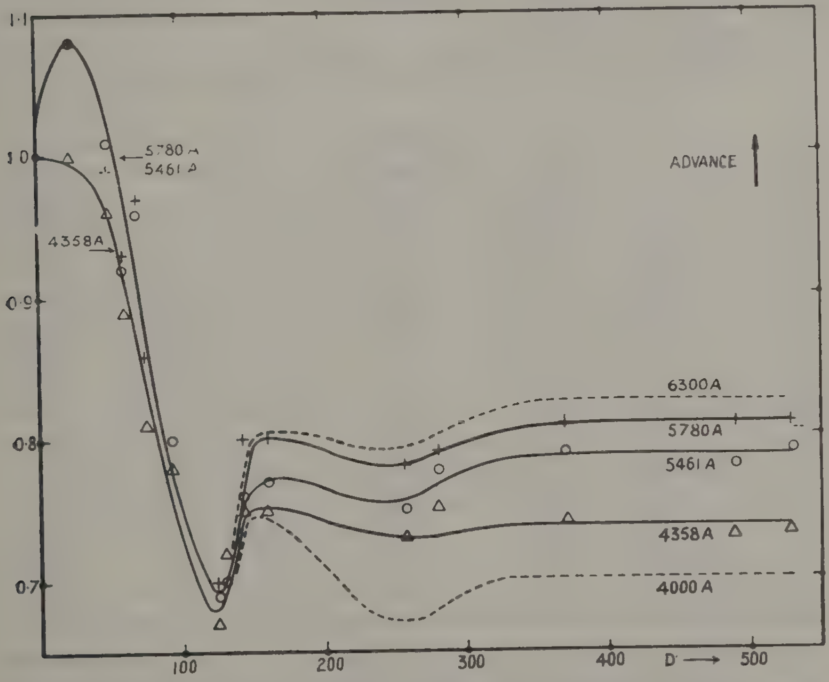
$$2(t'-t)=(dq+dp)\lambda=\lambda/2\pi\,.\,(\gamma-\gamma'), \quad . \quad . \quad . \quad . \quad . \quad (3)$$

in which dp is positive when the fringes in the silvered region are on the side of decreasing gap thickness relative to those in the unsilvered region. It is known from the Fresnel coefficients that the reflection phase change γ' at an air-dielectric interface is $\pm\pi$, the sign being in doubt. Following Rouard (1937) the negative sign is here accepted, the phase change being interpreted as an advance, *i.e.* the optical gap is smaller than the geometrical one.

If a film is of high reflectivity the reflection fringes are barely visible in the region where the silver film covers the rear surface. Transmission fringes can be used instead, the fringe contrast and sharpness being enhanced by coating the front surface with a highly reflecting, semi-transparent film. Equation (3) also holds for these transmission fringes which are unfortunately very diffuse (*cf.* fig. 3, Pl. XXXVI.).

Such measurements were made for the three mercury radiations 5780, 5461 and 4358 A. The phase changes, all of which are advances, are shown graphically in fig. 4 as fractions of π. For film thicknesses above 100 A. the error is $\pm 0{\cdot}02$, whilst for thinner films it reaches $\pm 0{\cdot}03$. With a given film the relative values for different wavelengths are more accurate ($\pm 0{\cdot}01$), since they are uninfluenced by uncertainties in the film thickness.

Fig. 4.



An alternative method of determining the air-silver phase change is to study the transmission fringes of equal chromatic order arising within an air-gap interferometer formed from two identically silvered optical flats Such a fringe pattern is shown in fig. 5 (Pl. XXXVI.) for 529A thick films, the mercury spectrum being superimposed upon one half of the field of view. The wave-number difference between any two transmitted fringes of order n and $(n+1)$ is

$$\Delta\nu=\nu_{n+1}-\nu_n=\frac{\lambda_n-\lambda_{n+1}}{\lambda_n\lambda_{n+1}}=\frac{1}{2t}[1-(\Omega_{n+1}-\Omega_n)] \quad . \quad . \quad . \quad (4)$$

the order increasing with decreasing wavelength. As the variation of Ω from order to order (*i. e.* with wavelength) is slight, the wave-number difference closely approaches the constant value $\frac{1}{2}t$. The wavelength difference $(\lambda_n-\lambda_{n+1})$ must therefore decrease rapidly as the wavelength shortens. The effective order of interference may be found the equation

$$n-\Omega=2t/\lambda_n=\frac{\lambda_{n+1}}{\lambda_n-\lambda_{n+1}} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (5)$$

in which, as N. Barakat has remarked, the fractional part represents Ω, n being of necessity integral. Here $\Omega=\gamma/\pi$ where γ is the desired air-silver reflection phase change.

The fringe pattern of fig. 5 (Pl. XXXVI.) is analysed in Table I. The wave-number differences are almost constant but, as seen from the fifth column, there is a gradual diminution due to the slowly varying phase change. The geometrical gap thickness can only be found from the approximate equation $\Delta\nu=\frac{1}{2}t$, and as a small error in t leads to a large percentage error in Ω, another procedure had to be adopted. The value chosen for t was such that the phase changes calculated from the fringes of equal chromatic order agreed as closely as possible with those determined from the Fizeau fringes. In this particular example the chosen value was $2t=1{\cdot}474\times10^{-3}$ cm. ($\Delta\nu=678$ cm.$^{-1}$), the calculated effective order being given in the sixth column. In the final column is the value of Ω obtained from Fizeau fringes. To ascribe a unique value to the geometrical gap implies an abrupt interface between the silver film and the air, and is therefore equivalent to defining the film thickness. In this way the phase changes at the two wavelengths 6300 A. and 4000 A. were found for films thicker than 125 A. These are represented by the dotted curves in fig. 4.

TABLE I.

Wave-length (A.U.)	Wave-number (cm.$^{-1}$)	Wave-length differences	Wave-number differences	6 . $\Delta\nu$	$n-\Omega$	Ω (Fizeau)
6457	15487	270	676	—	22·83	—
6187	16163	247	672	4037	23·82	—
5940	16835	229	675	—	24·82	—
5711	17510	211	672	4035	25·81	—0·81
5500	18182	196	672	—	26·80	—0·79
5304	18854	181	670	4033	27·79	—
5123	19524	171	674	—	28·77	—
4952	20198	160	670	4028	29·77	—
4792	20868	149	670	—	30·76	—
4643	21538	139	664	4020	31·75	—
4504	22202	132	671	—	32·72	—
4372	22873	125	673	4019	33·71	—0·72
4247	23546	—	—	—	34·71	—

An opaque film, whose phase change was required, was matched against a semi-transparent 370 A. film and the reflection fringes of equal chromatic order were studied (fig. 6, Pl. XXXVI.). It was known that the reflection minima coincided with the theoretical positions of the transmission maxima (*cf.* § 6), and the phase change could therefore be calculated from equations (4) and (5) in which it was possible to introduce an approximate value for the correction term $(\Omega_{n+1}-\Omega_n)$. Here, however, $\Omega=1/2\pi\,.\,(\alpha_2+\gamma)$ where α_2 is the already-determined phase change at the transparent film and γ is the desired phase change at the opaque film. The values obtained are close to those found for relatively thick semi-transparent films.

Wavelength (A.)	6300	5780	5460	4360	4000
γ (in terms of π)	0·84	0·82	0·80	0·73	0·70

Several conclusions may now be drawn from fig. 4.

(1) For thick films the air–silver reflection phase change increasingly departs from π with shortening wavelength, the difference in the phase change between 6300 A. and 5461 A. being almost the same as that existing between 4358 and 4000 A. These effects, corresponding to the more rapid falling off of the reflectivity as the wavelength is lowered, were also detected by Meggers (1921) for cathodically sputtered, thick silver films.

(2) As the film thickness is reduced the phase change oscillates, a well-marked drop occurring at about 125 A., *i. e.* for a reflectivity in green of about 25 per cent. Such a minimum was also noted by Rouard for a film thickness of 90 A. and a reflectivity of about 20 per cent. In this region the wavelength sensitivity of the phase change is very slight.

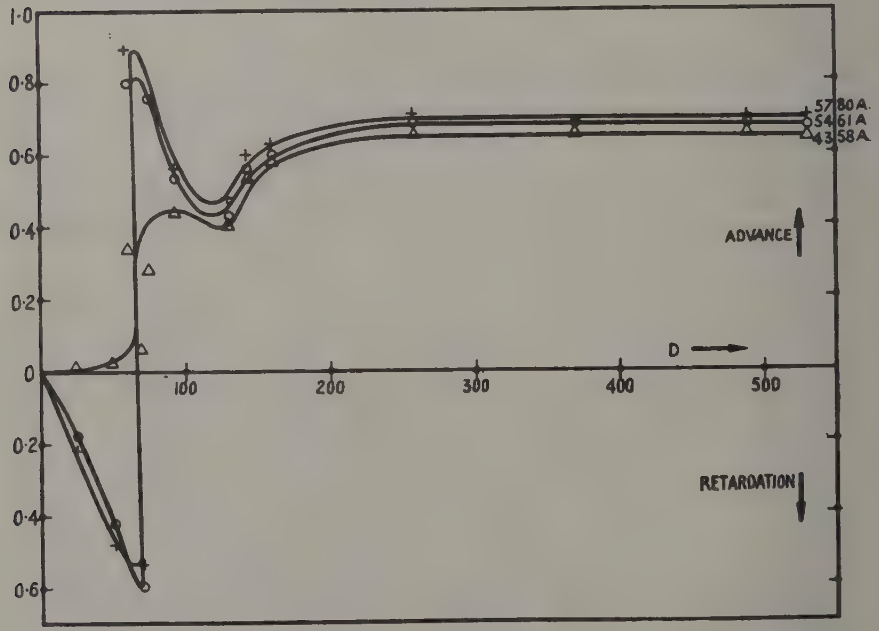
(3) For very thin films the phase change is close to π and even becomes greater than this quantity, thus being observed as an advance relative to that at a dielectric face. Such advances were also measured by Rouard for films below 32 A. thickness. With such thin films the phase change deviates least from π in the blue in accordance with the unusual behaviour of the air–silver reflection coefficient. Further reductions in the film thickness cause the phase change to attain the value π appropriate to the substrate alone.

§ 5. THE GLASS–SILVER REFLECTION PHASE CHANGE.

The glass–silver reflection phase change was measured by viewing reflection Fizeau fringes within a thin glass cover slip, the rear face of which was half-covered by the silver film. The phase change at the glass–air face, which according to theory is zero, was accepted as a standard. A knowledge of the film thickness was not required in order to calculate the glass–silver phase change which is provided by $2\pi\,.\,dp$, where dp is the observed fringe displacement at the boundary of the two halves of the interferometer (*cf.* equation (3)). When the film reflectivity was too great to yield satisfactory reflection fringes the front surface of the cover slip was coated with a highly reflecting, semi-transparent film and the transmission fringes were then studied. The lack of planeness of the faces of the glass slip often produced an irregular fringe spacing, and the very low glass–silver reflectivity of thin films resulted in a poor fringe visibility. For these reasons an accuracy better than ±0·04 cannot be claimed for the phase change measurements shown in fig. 7 as fractions of π.

For thick films the phase change is an advance which more closely approaches π the longer the wavelength. In the yellow and green a decrease in film thickness causes this advance to become a marked retardation, the transition, which covers a very narrow thickness range, occurring at that thickness (60 A.) for which the glass–silver reflection coefficient is at its minimum value. This retardation then diminishes, presumably reaching zero, the value appropriate to the uncoated dielectric. In the blue, on the other hand, the initial phase advance decreases and might become a very slight retardation before it finally vanishes. Similar results were found with the sputtered films studied by Rouard, the conversion of an advance into a retardation occurring in the neighbourhood of the glass–silver reflectivity minimum for wavelengths greater than 4358 A.

Fig. 7.



The reflection fringes obtained with the 61 A. film at the wavelength 5461 A. are shown in fig. 8 (Pl. XXXVI.), the fringes in the silvered part (the lower region of poorer contrast) being on the side of increasing gap thickness relative to those in the unsilvered part. The glass-silver phase change is therefore an advance. The film edge is not abrupt, and as the deposit becomes thinner this advance rapidly diminishes and even becomes a retardation, the fringes moving to the side of decreasing gap. This demonstrates the high sensitivity of the phase change to alterations in the thickness of such low reflectivity films.

The glass-silver phase change apparently deviates from π to a greater extent than the air-silver phase change for all wavelengths and film thicknesses, this behaviour probably corresponding to the relatively low reflectivity on the substrate side of the film.

§ 6. REFLECTION FRINGE ASYMMETRY.

The phase quantity χ, which is determined solely by the optical properties of the upper interferometer surface XX′ according to the expression

$$\chi=\beta_1+\beta_2-\alpha_1-\alpha_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

decides whether or not the reflection fringes shall be asymmetrical. It is, however, the combined properties of both interferometer plates which shapes the reflection fringe pattern and determines at what value of δ the reflection minima will occur. The theoretically predicted fringe behaviour is summarized in the following Table, in which it is to be noted that $\delta=2n\pi$ always corresponds to the transmission maxima and that $l=2t_1t_2\rho[\rho^2r_2(r_1r_2+t_1t_2)-r_1]$.

TABLE II.

χ	Condition	Reflection fringe pattern	Reflection minima in terms of δ
$(2N+1)\pi$	$l<0$	Symmetrical	$2n\pi$
$(2N+1)\pi$	$l=0$	Uniform field	—
$(2N+1)\pi$	$l>0$	Symmetrical	$(2n+1)$
$2N\pi$		Symmetrical	$(2n+1)\pi$
Between $(2N+1)\pi$ and $2N\pi$		Asymmetrical, minima closer to maxima of lower order	Between $2n\pi$ and $(2n+1)\pi$
Between $(2N+1)\pi$ and $(2N+2)\pi$		Asymmetrical, minima closer to maxima of higher order	Between $2n\pi$ and $(2n-1)\pi$

For relatively thick films the form of the reflection fringes was studied by simultaneously viewing the reflected and transmitted fringe systems. In this way the Fizeau fringes of fig. 9 (Pl. XXXVI.) were obtained with the 529 A. thick films at the wavelength 5461 A., faint secondary fringes arising from the 45° reflector being visible across the reflected pattern. The reflection minima agreed in position with the transmission maxima to within 0·02 of an order, a difference attributable to collimation and focusing uncertainties. Fringes of equal chromatic order were also observed with the same interferometer, the close agreement of the fringe positions throughout the visible spectrum being seen from fig. 10 (Pl. XXXVI.). This coincidence of the reflection minima with the transmission maxima was also discovered for all film thicknesses in excess of 142 A. (47 per cent reflectivity in the green), nor was any asymmetry perceptible. The phase term χ must, therefore, be close to an odd multiple of π for all wavelengths.

Because of the diffuseness of transmission fringes at low reflectivities a different method was adopted for thinner films. Only a half of each of the two glass interferometer plates was silvered and the interferometer was divided into four regions by the intersection of the two straight line edges of the films. In fig. 11 (Pl. XXXVII.) are seen the four separate Fizeau fringe systems observed in reflection (5461 A.) with the 93 A. films. Quoting the upper interferometer face first the four regions are glass/glass, glass/silver, silver/glass and silver/silver; the reflection minima in the two first-named regions coincide with the transmission maxima at $\delta=2n\pi$ and thus form reference marks. The fringe displacements for the 93 A. films are given below as fractions of an order, the fringes in the second region being on the side of decreasing gap thickness, *i.e.* decreasing δ.

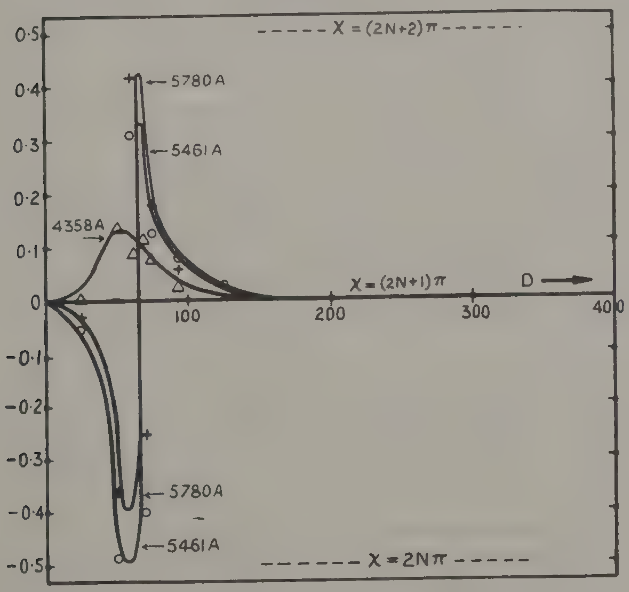
A.	glass/glass–glass/silver	0·067
B.	glass/glass–silver/glass	0·147
C.	silver/glass–silver/silver	0·067
D.	glass/silver–silver/silver	0·152
E.	glass/silver–silver/glass	0·080, 0·085.

The origin of the displacement A as the combined effect of the film thickness and the air–silver reflection phase change has already been described in § 3. The inequality of the displacements A and B shows that the addition of a silver film to the upper plate XX′ is not equivalent to adding one to the lower plate YY′, the difference E representing the influence of χ. A value for E can also be obtained from displacements C and D. As, however, the positions of the reflection minima are dependent upon the reflectivity of the lower interferometer face whenever χ deviates from a multiple of π, the displacement A might differ from C and displacement B from D; the two calculated values of E should, however, be identical. No conclusive evidence was found for such differences, this probably being due to the small alteration in the rear surface reflectivity (4 to 14 per cent) from one region to another.

In fig. 12 are seen the differences E expressed as fractions of an order, positive values signifying that χ is larger than $(2N+1)\pi$ and negative ones that it is below $(2N+1)\pi$. When E is $\pm 0{\cdot}5$, χ is an even multiple of π; this condition is closely approached with the 50 A. films in the green, the reflection fringes in the silver/glass and silver/silver regions then possessing the transmission-like appearance illustrated in fig. 13 (Pl. XXXVII.). When χ is no longer an integral value of π the reflected fringe pattern is asymmetrical in the sense indicated in Table II.

The 61 A. thick film is of interest In the yellow and the green the Fizeau fringe asymmetry is such that the minima are closer to the maxima of higher order. The displacement E, however, is apparently negative, not positive, since the minima in the silver/glass region appear to be on the side of greater order relative to those in the glass/silver region. Two explanations are possible for this behaviour : χ lies between $(2N+1)\pi$ and $(2N+2)\pi$ and some spurious edge effect gives the impression of a fringe movement towards larger orders instead of smaller ones, or χ is between $2N\pi$ and $(2N-1)\pi$, in which instance the displacements E in the yellow and green should read —0·58 and —0·69 instead of 0·42 and 0·31 The first explanation is favoured, the spurious influence being provided by the rapid variations in the glass-silver phase change which take place at the film edge (*cf*. fig. 8, Pl. XXXVI.), these variations bringing about marked changes in the value of χ in accordance with the arguments of the following section. In the blue, where the glass-silver phase change suffers only mild alterations and does not change sign, no edge effects are noticeable, and E is undoubtedly positive. This strong wavelength sensitivity of the film properties is illustrated in fig. 14 (Pl. XXXVII.), which shows fringes of equal chromatic order taken across the boundary

Fig. 12.



between the silver/silver and the glass/silver regions of an air-gap interferometer formed with the 61 A. films. In the blue the silver/silver minima are close to but on the low order (long wavelength) side of the glass/silver ones. As the wavelength increases the silver/silver minima become more displaced relative to those in the other region, but despite appearances this fringe displacement is towards longer and not shorter wavelengths, the strong slope of the silver/silver fringes at long wavelengths representing marked changes in the optical properties as the film thickness alters slightly.

§ 7. DISCUSSION.

If the film thickness is greater than about 150 A. the phase term χ is close to $(2N+1)\pi$ for all wavelengths, this being the value appropriate to the dielectric substrate. At the wavelength 4358 A. further reductions in film thickness cause a small initial increase in χ which is then followed by a decline, $(2N+1)\pi$ again being reached for the thinnest films. The behaviour at the two longer wavelengths is, however, different. At first χ increases rapidly, closely approaching $(2N+2)\pi$ in the yellow and possibly also in the green. Then at an approximate thickness of 60 A. it abruptly drops to about $2N\pi$; this rapid decline is followed by an increase, $(2N+1)\pi$ being reached from the lower side and not the upper one as in the blue. Holden (1949) who has utilized microphotometer traces to study the reflection fringe asymmetry in the green, discovered that the magnitude of χ rapidly diminishes from about $(2N+2)\pi$ to $2N\pi$ at an approximate air–silver reflectivity of 15 per cent, this being in agreement with the present work where the transition occurs at a reflectivity in the green of 12 per cent.

In equation (6) the quantity χ was defined as the sum of the air silver and glass-silver transmission phase changes minus the sum of the air–silver and glass–silver reflection phase changes. The two transmission phase changes are unknown and their contribution cannot be assessed. It is reasonable to assume that they suffer no rapid alteration in magnitude within the thickness range 50 to 70 A. In the same region the air–silver reflection phase change undergoes relatively mild oscillations, and the glass–silver reflection phase change must therefore exert the greatest influence. For example, at 60 A. this phase change abruptly increases by about $1{\cdot}4\pi$ in both the yellow and the green (phase advances are negative in sign). At the same thickness the fringe displacement E decreases by about 0·75 of an order, thereby indicating that χ has been reduced by an amount between π and 2π. In the blue, however, both the phase change and χ show only gentle variations and neither reverses sign. The general similarity between the curves in figs. 7 and 12 in the thickness range 0 to 100 A. is indeed striking, the two series of measurements being in harmony. Holden's suggestion that the reflection fringe asymmetry and the glass–silver reflection phase change are intimately related therefore receives experimental confirmation.

The experimental results may now be summarized and compared with the predictions of Table II.

(1) For thick and very thin silver films the phase term χ is close to an odd multiple of π, the light absorption is small and the ratio R/R' nearly unity. In an interferometer constructed of such films the reflection minima and transmission maxima are almost in perfect coincidence.

(2) In the immediate neighbourhood of the minimum in the glass-silver reflectivity curve the absorption and the ratio R/R' are large and χ deviates from $(2N+1)\pi$. The reflection fringes are then usually asymmetrical. In the blue, where the maximum absorption is relatively small, the deviation is less pronounced than in the green and yellow for which wavelengths χ closely approaches both $2N\pi$ and $(2N+2)\pi$ thereby producing nearly symmetrical reflection fringes, the minima of which coincide with the transmission maxima.

(3) In this same transition range χ should pass through the value $(2N+1)\pi$ in both the green and the yellow. As the ratio R/R' will then be very high, the reflection and transmission minima might coincide in an air-gap interferometer using such a film for the upper plate XX′; this is particularly likely when the reflectivity of the rear surface YY′ is high.

The optical properties of sputtered and evaporated silver films have been shown to possess similar characteristics. Gold films, whether they are sputtered, Rouard (1937) and Goos (1937), or evaporated, Sennett and Scott (1950), display the same general features as those of silver, and it is probable that these films will also produce reflection fringe asymmetry. Furthermore, the writer has observed asymmetrical reflection fringes using evaporated aluminium films, which Sennett and Scott have also shown to possess a peak absorption at a certain film thickness. The conclusions reached in this paper are therefore possibly typical of a number of metallic films.

Metal films have been regarded by Rouard and others as homogeneous layers of definable thickness and the principles of optical interference have therefore been applied to them. From the oscillating nature of the optical properties of the films it is then concluded that the optical constants vary with film thickness. A more satisfactory approach appears to be to adopt the ideas of Maxwell Garnett (1904, 1905), who regards the film as consisting of many regularly scattered globules, the optical constants of which are the same as those of the bulk metal. The sphere diameters, which are small relative to the wavelength of light, are not necessarily uniform, the characteristic of each film being the volume of metal spheres in unit volume of the film. Krautkrämer and Sennett and Scott have demonstrated that this theory provides a qualitative behaviour for many of the properties of metallic films, but that a number of features still remain unexplained. The theory might be improved by considering the non-spherical nature of the metallic particles.

ACKNOWLEDGMENTS.

The author wishes to thank Professor S. Tolansky for the continued interest he has shown in this work, which was made possible by the receipt of a grant from the Department of Scientific and Industrial Research. He also wishes to acknowledge the loan of the reflectometer by Mr. D. G. Avery and the valuable assistance given by the workshop staff of the Physics Department, Royal Holloway College.

REFERENCES.

AVERY, D. G., 1949, *Nature, Lond.*, **163,** 916.
GOOS, F., 1936, *Z. Phys.*, **100,** 95 ; 1937, *Ibid.*, **106,** 606.
HAMY, M., 1906, *J. Phys. Radium*, **5,** 789.
HOLDEN, J., 1949, *Proc. Phys. Soc.* B, **62,** 405.
KHAMSAVI, A., and DONALDSON, W. K., 1947, *Nature, Lond.*, **159,** 228.
KRAUTKRÄMER, J., 1938, *Ann. der Phys.*, **32,** 537.
GARNETT. J. C. M., 1904, *Trans. Roy. Soc.* A, **203,** 385 ; 1905, *Proc. Roy. Soc.* A, **76,** 370.
MEGGERS, W. F., 1921, *J.O.S.A.*, **5,** 308.
v. D. PAHLEN, E., 1912, *Ann der. Phys.*, **39,** 1567.
ROUARD, P., 1937, *Ann. de Physique*, **7,** 291.
SENNETT, R. S., and SCOTT, G. D., 1950, *J.O.S.A.*, **40,** 203.
STRONG, J., and DIBBLE, B., 1940, *J.O.S.A.*, **30,,** 431.
TOLANSKY, S., 1945, *Proc. Phys. Soc.*, **36,** 225.

FIG. 3.

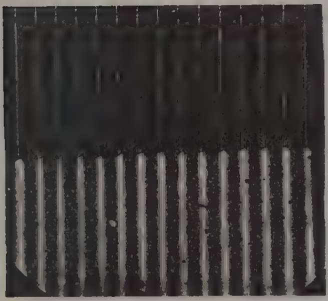


FIG. 5.

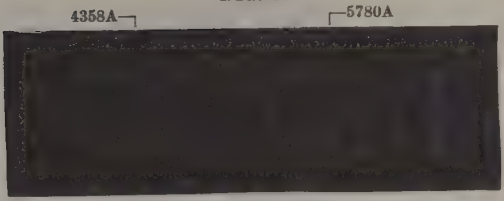

4358A
5780A

FIG. 6.

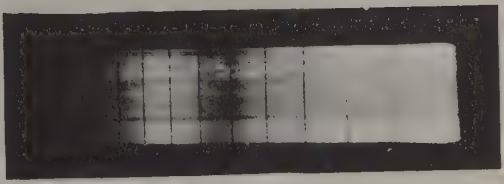


FIG. 8.

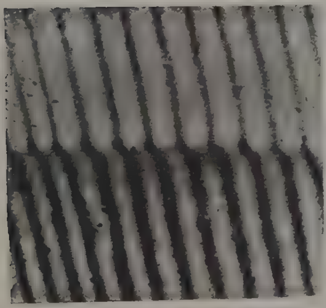


FIG. 9.

FIG. 10.

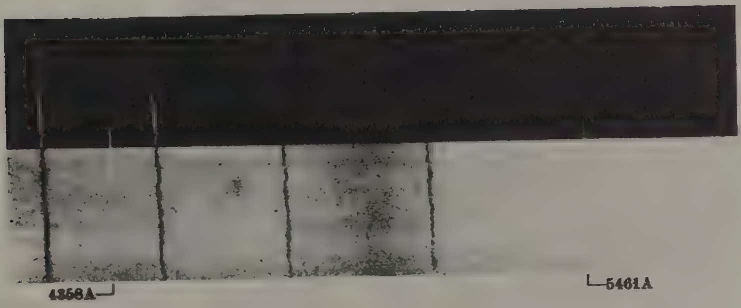

4358A
5461A

Fig. 11.

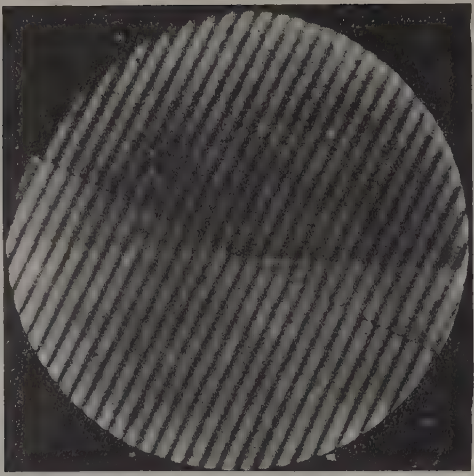

silver/glass
silver/silver
glass/glass
glass/silver

Fig. 13.

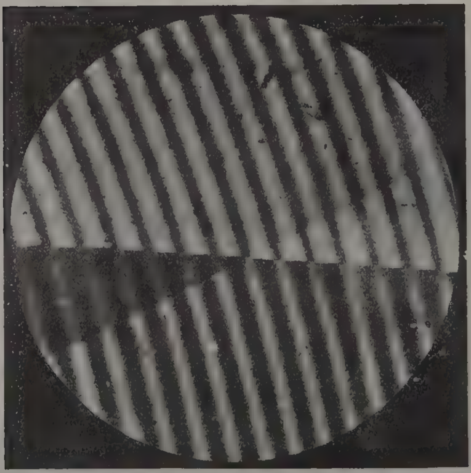
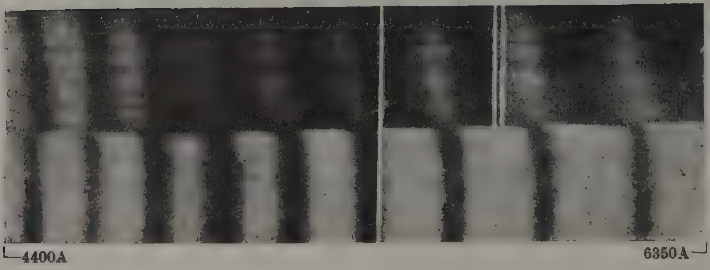

glass/glass
silver/glass
glass/silver
silver/silver

Fig. 14.


silver/silver
glass/silver
4400A
6350A

CXI. *The "Velocity" of Brownian Movement.*

By R. FÜRTH,
Birkbeck College, University of London[†].

[Received September 1, 1950.]

ABSTRACT.

The notion of the velocity of the random movement of a particle and the problem of its observation is discussed in view of certain incorrect statements in a recent paper by D. K. C. MacDonald.

THE problem of how to define and to measure the velocity of the random movement of a particle has frequently been discussed in the past. Recently D. K. C. MacDonald (1950) has contributed to the problem by applying to it the modern notion of the "frequency spectrum" of a random movement and by deriving an expression for the "observable" mean square fluctuation $\overline{\delta_v^2}$ of the velocity which he attempts to test on some early observations of the Brownian movement of small particles by F. Exner (1900). However, his analysis rather confuses the real issue and the application to the observations of Exner is partly erroneous. I, therefore, wish to make the following brief remarks on the problem.

(1) Let x be a "unidimensional stochastic function" of time t. This means that a function $x(t)$ does not exist in the usual mathematical sense, but that there exists a probability density function $p(x,t)$ for observing a certain value x at time t. Consequently all moments x^s are ordinary functions of time and their derivatives exist. In particular, $V \equiv d(\bar{x})/dt$ is the so-called drift velocity (which may be zero), and the quantity $D \equiv \frac{1}{2}d(\overline{x^2})/dt$, which, in general, is different from zero, is a convenient measure for the rate of fluctuations of x (D can, of course, be independent of time). But the quantity dx/dt does *not* exist at all.

If in particular x is a coordinate, a corresponding velocity component in the strict sense of the word cannot be defined. *Vice versa*, if a velocity component v is primarily introduced as a stochastic variable the corresponding position co-ordinate has no meaning. This is the well known situation in quantum mechanics where either coordinates or momenta of particles can be chosen as representing variables, but not both at the same time. It leads directly to the uncertainty principle for the simultaneous measurement of position and velocity.

(2) There is, of course, no justification for regarding a physical quantity x as a stochastic variable under conditions where quantum effects can be neglected. For the random character of the variation of

---

[†] Communicated by the Author.

x in such a case is only due to the influence of " random external disturbances " on the observed system which are supposed to be not accessible to direct measurement. Thus x can be regarded as an ordinary function of time for an individual observed system, and the system, therefore, possesses at each moment a well defined velocity v which, in the absence of a drift velocity, will irregularly fluctuate in time.

One can, evidently, also define a " velocity " v^* by the difference quotient

$$v^* = \frac{\Delta x}{\Delta t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

over a finite time interval Δt. This v^* will be found to be practically identical with v as long as Δt is smaller than a certain time interval τ which depends on the mechanism of the observed system and which may appropriately be called its " characteristic time "(†).

If $\Delta t < \tau$, the statistical characteristics of v^* must be identical with those of v. If, on the other hand, Δt is considerably larger than τ the statistical characteristics of v^* may be very different from those of v. In any case, however, a statistical analysis of the behaviour of v^* is absolutely equivalent to a statistical analysis of a record of displacements Δx at equal time intervals Δt, and so no useful purpose is served by the introduction of the quantity v^*.

(3) So far we have assumed that x is measured by means of an " ideal " instrument. In actual fact, the result of the observation by a real measuring device will usually be in the form of a graph representing the dependence of a " reading " y on time, y being a suitable measure of x. The rate of change w of y is obtained at any point of this graph by constructing the tangent to the arc of the curve at that point. This, of course, is only psssible with a finite accuracy, depending on the " resolving power " δy of the instrument.

The problem of the accuracy of the simultaneous determination of y and w has been discussed in detail by the author (1935, 1948), and he has shown that it is limited by a relation which is analogous to the quantum mechanical uncertainty principle.

Within the limit δy of accuracy, one can also measure a quantity w^* defined by

$$w^* = \frac{\Delta y}{\Delta t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where Δt is again a finite time interval. As a rule, it will be found that w^* is identical with w as long as Δt is smaller than a certain time interval δt which we may call the " resolving power in time " of the instrument

---

(†) τ need not necessarily be the same as the so-called " relaxation time " which plays a decisive part in MacDonald's considerations ; this latter quantity is only defined for phenomena which are the outcome of the opposing influences of inertial and dissipating processes.

as the record can clearly give no indication whatsoever about the variation of the measured quantity x within time intervals smaller than δt(†). It has to be borne in mind, however, that both δt and δy cannot be defined precisely and that their values are necessarily arbitrary within considerable margins.

If one can make δt of the order of magnitude τ (or smaller) the record can be used for investigating the statistical properties of v. But if, as usual, δt is much larger than τ the record can only be used for the investigation of the statistical properties of Δx for time intervals Δt larger than δt. This holds *a fortiori* for discontinuous observations of y at equal time intervals Δt. Clearly no complicated mathematical analysis is required to come to this almost trivial conclusion, and no analysis can reveal anything beyond the above statement.

(4) Let us finally consider the practical example used by MacDonald to test his theoretical relations. The observations of Exner consisted in following, by means of a camera lucida, the movemnt of a Brownian particle as closely as possible with a stylus on a blackened glass plate. This curve was subsequently projected on a screen and the total length of the curve measured. The " velocity " of the Brownian movement was then taken to be this length divided by the total time taken for the observation.

Evidently in this case δt is the " reaction time " of the observer for this sort of observation which is large compared with τ. Thus, according to (2), the measured " velocity " is simply $\overline{(\delta s)}/\delta t$ where $\overline{(\delta s)}$ is the average arc of the curve described in time δt.

As the observed Brownian motion is a two-dimensional projection of the actual three-dimensional path, one finds easily for the probability density $p(\delta s, t)$ of δs in time t

$$p(\delta s, t) = \frac{\delta s}{2\mathrm{D}t} \exp\left(-\frac{(\delta s)^2}{4\mathrm{D}t}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where D is the diffusion coefficient of a suspension of identical particles. From (3) one obtains at once

$$\overline{(\delta s)} = \int_0^\infty p(\delta s, t)\delta s d(\delta s) = \sqrt{(\pi \mathrm{D}t)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Thus finally

$$w^* = \sqrt{(\pi \mathrm{D}/\delta t)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

MacDonald, however, identifies his $\sqrt{\overline{\delta_v^2}}$ with w^* which is incorrect, as his $\overline{\delta_v^2}$ refers to a one-dimensional Brownian movement ; he also

---

(†) δt may have the character of a relaxation time but not necessarily so ; *e.g.* if the record is obtained by trying to follow the movements of a spot on a projection screen with a pencil, δt will mainly be determined by the time the nervous impulses take to travel from the eye to the brain and back to the muscles of the hand.

mixes up a mean absolute value with a root mean square value, a procedure he himself strongly criticises in his paper.

For spherical particles of radius a

$$D=\frac{kT}{6\pi\eta a}. \quad \quad (6)$$

From (5) and (6) an expression for w^* is obtained which, of course, is of the same order of magnitude as MacDonald's expression (16), but there is no justification for his different numerical factor, although it is allegedly obtained by a refined mathematical analysis.

From the values of w^* given in Exner's paper for particles of various sizes, one obtains from (5) and (6) values of δt between 0·1 and 0·2 sec. which are of the order of magnitude to be expected. No other conclusions can be drawn from these experiments.

REFERENCES.

Exner, F., 1900, *Ann. d. Phys. Lpz.*, **2**, 843.
Fürth, R., 1935, *Zs. f. Phys.*, **81**, 143.
Fürth, R., 1948, *Science Progress*, **143**, 396.
MacDonald, D. K. C., 1950, *Phil. Mag.*, **41**, 814.

CXII. *The Use of Atmospherics to Study the Propagation of Very Long Radio Waves.*

By F. F. Gardner,
Cavendish Laboratory, Cambridge *.

[Received October 5, 1950.]

Summary.

The "level" of integrated atmospherics received on narrow-band receivers tuned to a series of different frequencies between 3·5 and 50 kc./s. has been recorded. The band-width of the receivers was about 200 c./s. and the time of integration about one minute. The results were analysed so as to eliminate the unknown spectral distribution of energy at the source, and so as to show up the way in which propagation conditions changed with distance and with time of day. The results are shown in fig. 4, from which the most significant effect is seen to be a marked attenuation of waves of frequency below 10 kc./s. when travelling over distances of 1000 km. or more. This attenuation is discussed and is related to the quasi-sinusoidal waveform often assumed by atmospherics which have travelled great distances.

A decrease in the level of atmospherics has been found to occur during sudden ionosphere disturbances (S.I.D.'s) on frequencies below about 10 kc./s. This contrasts with the known increase on frequencies above about 20 kc./s.

§1. Introduction.

A study of the propagation of the very longest radio waves entails the use of powerful sending stations, using large ærials, since the difficulty of radiating appreciable power increases as the wavelength is increased. For this reason the senders which have been used are those available for commercial radio services, and the longest waves at present available in this way have a length of about 18 km. (16 kc./s.). Much useful information has been gained about the ionosphere by studying the waves emitted by these stations (Best, Ratcliffe and Wilkes 1936, Budden, Ratcliffe and Wilkes 1939) and it is desirable that observations should be made on even longer wavelengths. The construction of special experimental senders for this purpose would be a large and expensive undertaking and an attempt has therefore been made to use instead naturally occurring atmospherics as sources of very long waves. This paper gives some of the results obtained at Cambridge in this way during the years 1948 and 1949.

It is generally agreed that radio atmospherics have their origin in the electrical discharges associated with thunderclouds. The discharges, being impulsive in nature, radiate over a wide frequency range and can be used as senders for investigating the propagation of very long waves. The large energy available in atmospherics at these wavelengths renders them easily observable compared with man-made and circuit noise. One of the major difficulties encountered in the use of atmospherics in this way is that the " waveform " of the original discharge is not known, so that the power radiated on different frequencies by any one atmospheric is not known, and the experiments must be planned so that the results are either independent of, or give us information about, the distribution of radiated power amongst the different frequencies.

---

* Communicated by J. A. Ratcliffe.

Another difficulty is that atmospherics are produced at different distances from the receiving point, and without some knowledge of, or assumptions about, their location, it is impossible to make useful deductions about the propagation of the associated waves. Fortunately we have been able, thanks to the assistance of the Director of the Meteorological Office, to make use of the chain of widely-spaced radio direction-finders which the Meteorological Office use to locate lightning flashes in their so-called " Sferics " organization. In this way it was known when fairly localized storm-centres were in action at distances less than about 1500 km.

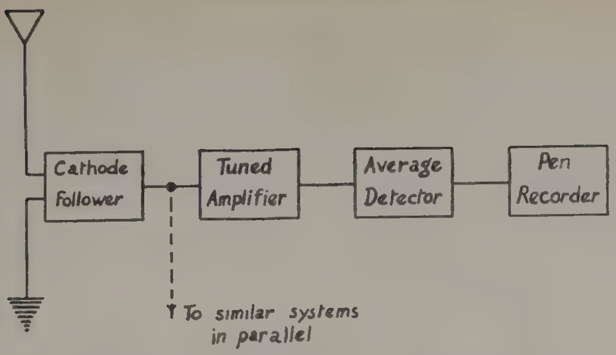
The method employed was to determine four points on the " spectrum " of the integrated received atmospherics, by recording the level on four different frequencies, and then to investigate how this spectrum changed from day to night, and with the distance of the storm centre. The assumption was then made that changes in the spectrum corresponded to changes in the propagation conditions, brought about either by the change from day to night, or by the change in the distance of the storm.

§2. METHOD OF OBSERVATION.

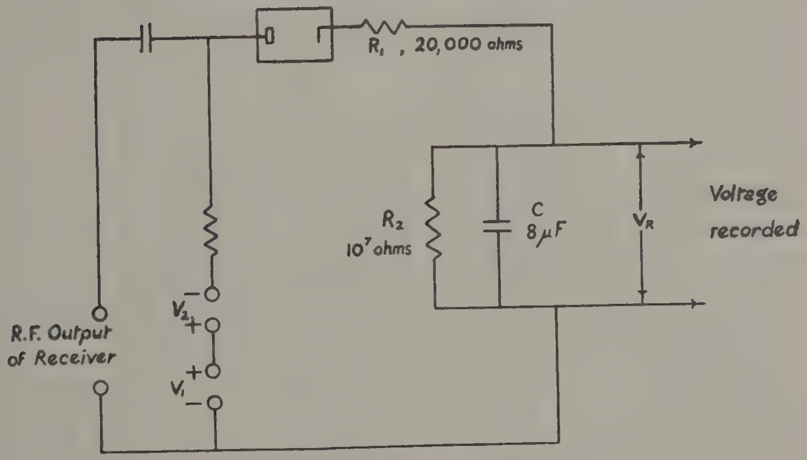
The results to be described were obtained from records of the average field strength of atmospherics on a number of different frequencies. The equipment used on one frequency is shown schematically in fig. 1 (*a*). A vertical ærial was connected to the common input of the various tuned receivers. Each receiver selected the appropriate frequency component from the atmospherics arriving. Its output was then rectified and averaged with a time constant of about one minute. The resulting voltage was recorded on paper moving at one inch per hour.

The receivers were all of similar type. Each consisted of a tuned input stage followed by two tuned amplifier stages, and had an overall bandwidth between half-power points of approximately 200 c./s. The radio frequency output fed the detector circuit and was also available for viewing on an oscilloscope. The detector circuit is shown schematically in fig. 1 (*b*). The time constant for discharge of the condenser C was about 80 seconds which was very long compared with the average interval between the arrival of atmospherics. Hence a fairly steady output was obtained which represented a running mean of the input signal. The voltage V_R across the condenser C at any instant therefore depended on the signals received

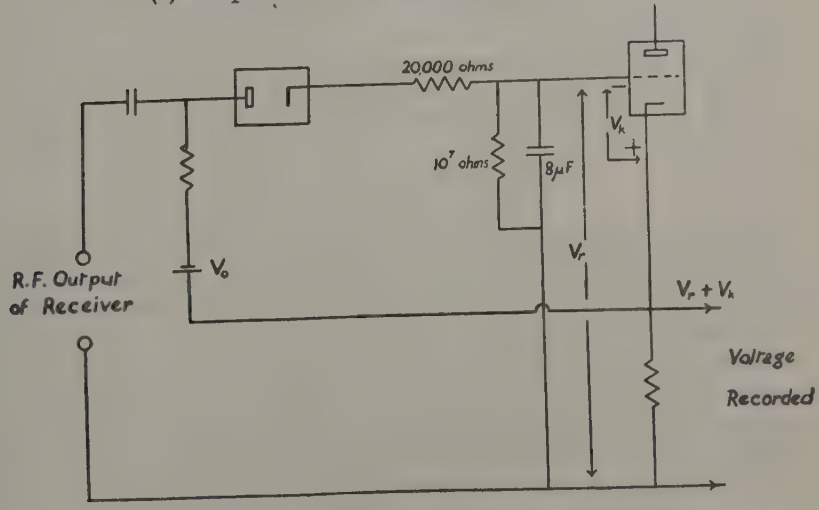
Fig. 1. (*a-c*)


Cathode Follower
Tuned Amplifier
Average Detector
Pen Recorder
To similar systems in parallel

(*a*) Arrangement of atmospherics receiver.


R_1 , 20,000 ohms
R_2 10^7 ohms
C 8μF
V_R
Voltage recorded
R.F. Output of Receiver
V_2
V_1
+
−

(*b*) Simplified form of detector circuit.


20,000 ohms
10^7 ohms
8μF
V_k
+
V_r
V_0
$V_r + V_k$
Voltage Recorded
R.F. Output of Receiver

(c) Actual form of detector circuit.

during a time of the order of 80 seconds. In the simple circuit of fig. 1 (*b*) this voltage would affect the bias on the detector diode. To prevent this a voltage V_1 equal to V_R was connected in series with the diode. This was actually achieved, as shown in fig. 1 (*c*), by applying the voltage V_R to a cathode follower valve, and using the output of this valve to supply bias to the diode. The cathode voltage of a cathode follower exceeds the grid voltage by a small amount V_K which is roughly constant, so that the output voltage of the cathode follower was V_R+V_K. The excess voltage V_K was compensated as explained below.

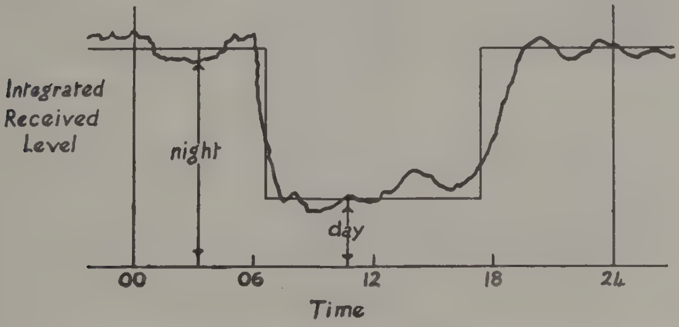
When the radio frequency output was examined on an oscilloscope, it was found to consist of atmospherics of short duration and fairly large amplitude and of continuous " noise " of small amplitude, much of which was known to be due to " man-made " interference. On some frequencies, also, a small output due to signals from radio stations was sometimes observed. To reduce the effect of these unwanted voltages, an additional small voltage, V_2 in fig. 1 (*b*), was connected in series with the diode, which prevented it from conducting until the level of the signal output from the receiver exceeded a small threshold about equal to V_2. The voltage V_2 was actually derived, as shown in fig. 1 (*c*), from a constant voltage source V_0 which served also to compensate for the voltage V_K in the output of the cathode follower, so that $V_2=V_0-V_K$. V_2 was equal to 2 or 3 volts, which greatly reduced the contributions from the unwanted signals without appreciably affecting the wanted atmospherics.

The apparatus was set up at the Radio Field Laboratory in Cambridge, where the general level of man-made noise was known to be low. Records were made between February 1948 and December 1949, at first on two frequencies, and after June 1949 on four frequencies. The frequencies were selected from time to time from amongst 3·5, 5, 7·5, 10, 14, 22, 27, 33, 40 and 48 kc./s. The receivers were calibrated from time to time, at first by means of a C.W. signal generator, and later by means of a special impulse generator which produced a regular succession of pulses each of duration 1, 4 or 14 microseconds. These calibrations were only used to check the linearity of the receivers, and in the deductions to be described no use is made of the relative calibrations on the different frequencies.

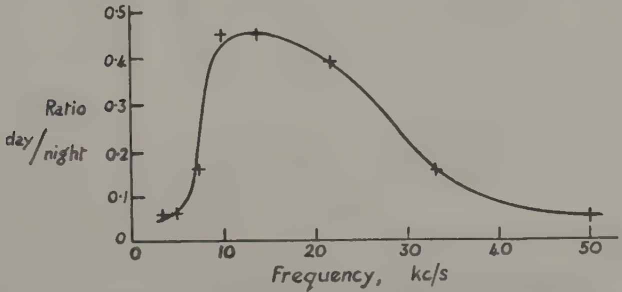
§3. RESULTS OBTAINED.

The records were analysed so as to bring out effects which might be produced by changes in the conditions of propagation. For this purpose it was assumed, as a first approximation, that waves of all frequencies were propagated without appreciable attenuation from storm centres which were comparatively near (less than about 600 km.). It was next assumed that, in the absence of near storms, the atmospherics represented sources at a considerable distance (greater than 1000 km.) and that the recorded integrated levels measured relative to their values at the source would vary with frequency in a way which depended both on the nature of the source and on the propagation characteristics. The effect of the propagation characteristic at night was established by comparing the normal night-time level with the level produced by near storms, as explained in sub-section (b) below. If, next, propagation conditions change as we pass from night to day it should be possible to establish the change by comparing, in the absence of near storms, the levels recorded by day and by night at the different frequencies. This comparison is made in sub-section (a) below. It is convenient to consider these two comparisons in the reverse order.

Fig. 2. (*a*-*b*)



(*a*) Typical atmospherics record in the absence of near storms.

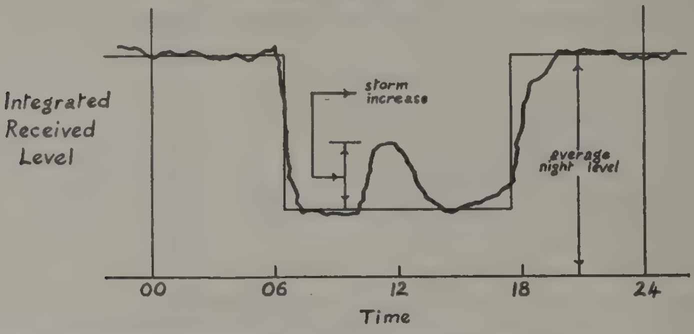


(*b*) Ratio, day level to night level as a function of frequency.

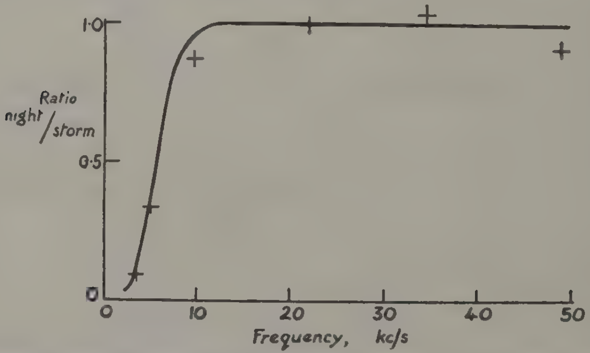
(a) *Change from day to night.*

Fig. 2 (*a*) shows a typical record for a 24-hour period when it was known that there was no storm centre within 1000 km. of Cambridge. From a record of this type it is possible to estimate, with some significance, the average level in the few hours before sunrise and the level in the few hours after sunrise, and to determine their ratio. The average value of this ratio, deduced from a large number of records obtained on different frequencies, is exhibited as a function of frequency in fig. 2 (*b*). Often local storm centres would develop during the afternoon so as to complicate the records, but by determining the daytime level from the few hours just after sunrise, and at the same time disregarding the spasmodic increases produced by isolated near discharges, a significant ratio was obtainable. Fig. 2 (*b*) appears to indicate that the onset of daylight conditions represents a marked increase of attenuation for frequencies less than 10 kc./s. and greater than 20 kc./s. but near 10 kc./s. there is little extra attenuation.

Fig. 3. (*a*-*b*)



(*a*) Atmospherics record, showing effect of near storm.



(*b*) Ratio, night level to increase due to storm, as a function of frequency.

(b) *Comparison of near and distant storms.*

Additional evidence is derivable from the increase in the average field strength which occurs when a comparatively close storm centre becomes active. The type of record then obtained is shown in fig. 3 (*a*). On several occasions records of this type were obtained, and it could be seen from the " Sferics " charts that a comparatively local storm centre had become active at some distance less than 1000 km. Under these conditions it was possible to estimate the increase in average field strength on the frequencies being recorded. The results were studied by computing the ratio (average night field strength) ÷ (increase due to storm). The absolute value of the ratio depends, of course, on the intensity of the particular storm being considered ; what is of interest, however, is not the absolute magnitude of the ratio but the way it changes with frequency.

For any one storm we have therefore normalized the ratios obtained on the different frequencies so that the ratio for a frequency of 22 kc./s. is put equal to unity. During the short period of recording on 3·5 kc./s., only three storms of short duration and completely localized within 600 km. were available, and these provided the mean " night-to-storm " ratios for 3·5, 5 and 22 kc./s. shown in fig. 3 (*b*). Other storms, more extensive in range or duration, yielded values showing the same general trend at the low frequency end, while records made at different times on other frequencies yielded the points shown on the curve for frequencies between 10 and 50 kc./s. In interpreting these results it should be remembered that the field on 3·5 kc./s. is predominantly the radiation field at distances greater than about 20 km. and all the storms considered were further away than that.

§4. Discussion of Results.

We start our discussion of the results by considering the implications of fig. 3 (*b*). We first make the assumption, previously mentioned, that all frequencies are propagated equally well from a near storm centre, and that the normal night-time level represents storms at a great distance. Then fig. 3 (*b*) gives a measure of the relative attenuation suffered by waves of different frequency as they travel over great distances at night. The results are normalized so that the attenuation factor for a frequency of 22 kc./s. is taken as unity. Of course the distances of the sources at night are very indefinite, but the results appear to show that, when the sources are at a considerable distance (greater than 1000 km.), the lower frequencies are much more heavily attenuated than the higher and that the " low-frequency cut-off " occurs just below 10 kc./s. Although this conclusion rests mainly on observatious made on a few occasions only, we consider it to be quite sound because it is supported by observations made on several other occasions when different frequencies were in use, or when the storm centre was not accurately located.

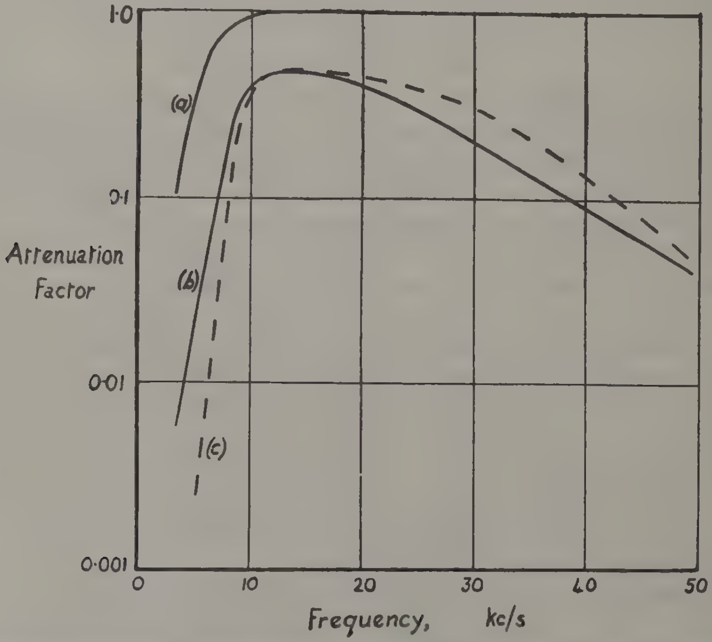
We next consider the curve of fig. 2 (*b*) which shows the extra attenuation introduced on the different frequencies as we pass from night-time to daytime. Since this curve refers to conditions when there were no storms within 1000 km. we assume that it represents propagation over great distances. Fig. 2 (*b*) is the result of a preliminary analysis covering the midsummer months of 1949 (more than 60 days' observations are involved in the determination of the value for 5 kc./s.), and we consider that it represents a well-established result and shows that the incidence of daylight has the effect of producing a considerable increase in the attenuation of waves propagated over great distances on frequencies below about 10 kc./s.

Fig. 4 (*a*) shows the attenuation curve of fig. 3 (*b*) for night, re-plotted on a logarithmic basis. Fig. 4 (*b*) represents the product of the curves of figs. 2 (*b*) and 3 (*b*), plotted in a similar way, and therefore represents the attenuation curve for day.

Although there may be some doubt about the absolute magnitudes of the attenuation factors shown in this figure, we feel reasonably confident that it indicates, in a general way, how the attenuation factor depends on frequency for transmission over distances of 1000 km. or more. We shall here concern ourselves with the results represented by the portion of the curve below 10 kc./s.

In considering these results it is convenient to think of the space between the ground and the ionosphere as forming a four-terminal network, with the input terminals at the lightning flash and the output at the receiving ærial. Our experimental results provide a description of the overall frequency characteristics of this network and, in particular, show that there is a sharp cut-off at frequencies below about 10 kc./s., but do not indicate how these characteristics come about. It might, for example, result from the combination of separate ground- and sky-waves. In this connection, however, one point of significance emerges. The attenuation of a ground wave of frequency 5 kc./s. by absorption and diffraction, when travelling over a distance of 2000 km., can be calculated and the attenuation factor is found to be about 0·3. But the daytime attenuation factor measured on this frequency is much lower, about 0·02, as shown in fig. 4 (*b*). We deduce that the strength of the daytime field is much less than that of

Fig. 4.



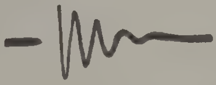
Attenuation factor as a function of frequency for (*a*) night, (*b*) day, (*c*) day during a Sudden Ionospheric Disturbance.

the ground-wave alone. In explaining this result it will probably be necessary to remember that the wavelength (60 km.) at this frequency is comparable with the distance between the ground and the " reflecting " region, so that it is doubtful whether we are entitled to think of the problem in terms of separate ground- and sky-waves.

It is not the purpose of this paper to discuss the explanation of the results here presented, but I understand that Dr. K. G. Budden is preparing a paper in which he suggests a theoretical explanation.

Before leaving the discussion of the curves of fig. 4 it is of interest to relate them to the waveform of an atmospheric received on an aperiodic amplifier. If we assume that the atmospheric is a sharp impulse lasting considerably less than 2×10^{-4} sec. (the time period corresponding to a frequency of 5 kc./s.), then simple transient analysis tells us what output we should expect if it is applied to a four-terminal network having the transmission characteristic of fig. 4. The result would be a quasi-sinusoidal oscillation with " period " roughly equal to the cut-off frequency of the network.

Fig. 5.



Waveform of atmospheric from a distant source.

Now it is well-known (Appleton and Chapman 1937) that atmospherics received from the greater distances very frequently show a smooth quasi-sinusoidal waveform with a period of the order of 150 microseconds. An example of one of these waveforms is shown in fig. 5 and it is clear that since it is so smooth and regular it would be difficult to interpret it in terms of a series of separate reflections from the ionosphere. It is, however, very much what we should expect as the result of applying an impulse to a transmission system with an attenuation curve as shown in fig. 4, and it seems probable that, once this attenuation curve can be explained, the origin of the quasi-sinusoidal atmospheric waveforms will be clear.

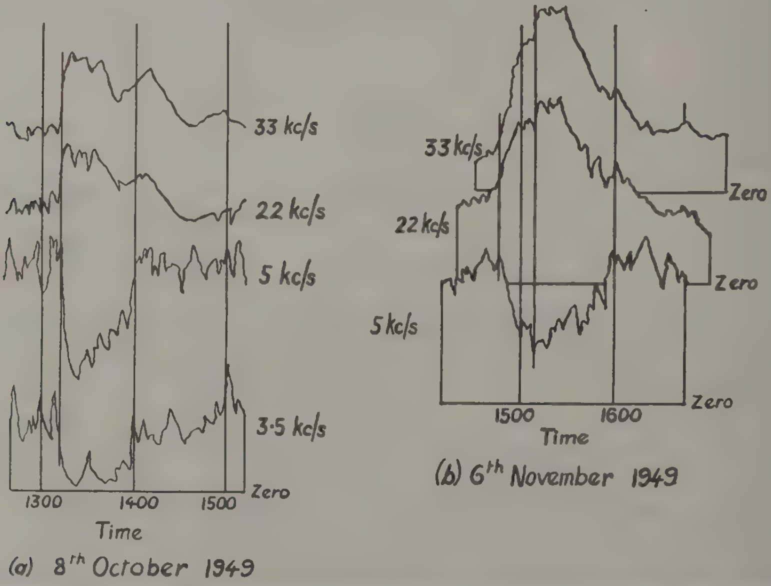
§5. Effect of Sudden Ionosphere Disturbances (S.I.D.'s).

Bureau (1937) has shown that the Sudden Ionospheric Disturbances (S.I.D.'s) of the type associated with solar flares, and frequently with " Dellinger Fadeouts ", very often coincide with an increase of the average field strength of atmospherics recorded on frequencies in the range 20–40 kc./s. We have confirmed his observations in this range, but our records on the lowest frequencies have revealed the unexpected phenomenon that there the average field strength often *decreases* during the S.I.D. This effect has been observed most frequently on 5 kc./s. Examples are shown in fig. 6.

We have already seen that, during the day, the average field strength on the lowest frequencies is very small, unless there is a fairly near storm, and indeed if there are no storm centres within 2000 km. the average field strength recorded on 5 kc./s. is so low that a further decrease due to an S.I.D., if it does occur, is not noticeable. For a frequency of 3·5 kc./s. the corresponding distance is about 1400 km.

If the storm centre is at distances between 1500 and 2000 km., it has been observed that S.I.D.'s can reduce the mean field strength by a factor of 10 : 1 on a frequency of 5 kc./s., and for distances between 600 and 1500 km. a reduction of 3 : 1 has been observed. Observations on 3·5 kc./s. have been less extensive, but on one occasion the field due to a storm centre at a distance of 800–1500 km. was reduced to one fifth during an S.I.D.

Fig. 6.



Records of integrated atmospherics at times of Sudden Ionospheric Disturbances.
(*a*) 8th October, 1949 (zero level shown only for 3·5 kc./s.);
(*b*) 6th November, 1949.

Although the effects associated with S.I.D.'s have depended on the distance of the storm centres, we have attempted to represent them in a general way by the curve (*c*) in fig. 4. In examining this curve it is interesting to notice the different effect of an S.I.D. on the resultant signal produced at a great distance on different frequencies. For frequencies below 10 kc./s. the signal decreases : for frequencies between 10 kc./s. and about 200 kc./s. the signal increases : for frequencies above about 1 Mc./s. the signal decreases (Dellinges effect). A satisfactory theory of propagation will have to explain these results.

§6. ACKNOWLEDGMENTS.

This work was done as part of a programme of radio research supported by the Department of Scientific and Industrial Research at the Cavendish Laboratory, Cambridge. I wish to thank Mr. J. A. Ratcliffe for much advice and encouragement during the experiments and assistance with the writing of this paper. I am indebted also to the Commonwealth Scientific and Industrial Research Organization of Australia, to Sidney Sussex College, Cambridge, and to the Director of the Cavendish Laboratory for grants which enabled me to work in Cambridge.

REFERENCES.

APPLETON and CHAPMAN, 1937, *Proc. Roy. Soc.* A, **158**, 1.
BEST, RATCLIFFE and WILKES, 1936, *Proc. Roy. Soc.* A, **156**, 614.
BUDDEN, RATCLIFFE and WILKES, 1939, *Proc. Roy. Soc.* A, **171**, 188.
BUREAU, 1937, *Nature, Lond.*, **139**, 110.

CXIII. *The Force Acting on a Body Moving Uniformly through a Gas containing Sound Waves.*

By F. R. N. Nabarro,
Department of Metallurgy, The University of Birmingham *.

[Received September 25, 1950.]

Summary.

When a body travels through a gas containing an isotropic distribution of sound waves it suffers a retarding force proportional to its area, to the energy density of the sound, and to its velocity. Three effects might contribute to this force: (*a*) the distribution of sound with respect to the moving body is no longer isotropic; (*b*) the apparent wavelength of the sound is altered by the motion of the body; (*c*) the non-linear terms in the equation of motion lead to a coupling between the sound waves and the steady flow of gas round the body. Effect (*c*) is shown to produce no force. Effects (*a*) and (*b*) are evaluated in the case of a sphere, and the force is shown to be expressible in terms of the force exerted by a single sound wave on a stationary sphere, already calculated by L. V. King.

§ 1. Introduction.

It is well known that a body moving uniformly through a non-viscous fluid at rest is not acted on by any force. The resultant couple also vanishes if the body is symmetrical (Lamb 1932). If the body is at rest, and a plane sound wave impinges on it, the sound wave is scattered, and a force is exerted on the body. This force has been calculated by Rayleigh (1905) in the case where the obstacle is the flat end of a cylinder containing the gas, and by King (1934) for the case of a spherical obstacle. (King actually treats the more general problem of a free sphere of density ρ. The result for a fixed sphere is obtained by taking ρ infinite.) An apparent discrepancy between their results is discussed in § 2 of the present paper.

If a symmetrical body lies at rest in a gas containing an isotropic distribution of sound waves, the resultant force on it clearly vanishes. If it moves with velocity v, it experiences a retarding force proportional to its area, to the energy density of the sound waves, and to v/c, where c is the speed of sound. In § 3 a plausible argument is advanced for believing that this force should vanish. The exposure of this fallacy illustrates an unexpected difficulty which arises if the thermal motion of the molecules of a gas is analysed into Fourier components in the way customarily used for solids. The result to be obtained is proportional

---

* Communicated by the Author.

both to the speed of the body and to the energy density of the sound waves, and therefore involves terms of the third order of small quantities in the equations of motion. In § 4 the equations are developed to the necessary order. In § 5 the explicit interaction between the uniform motion of the body and a single sound wave is evaluated and found to lead to no resultant force. However, an isotropic distribution of sound waves in the gas is not isotropic in axes attached to the moving body, and in § 6 it is shown that this anistropy leads to a force of the kind expected.

This problem has a certain intrinsic interest. I was actually led to study it by reading a paper by Leibfried (1950) in which he applies similar considerations to estimate the resistance to the motion of a dislocation in a crystal caused by its scattering of sound waves in that solid. His result for this more complicated problem depends upon several drastic simplifying assumptions. He finds a resistance so large that the terminal velocity of a dislocation in a crystal under a reasonable stress could never become comparable with that of sound. Since this result would render impossible a type of process which plays an important rôle in some modern theories of plastic deformation (Frank 1948), it seemed worth while to treat this related, but simpler, problem in detail. The results are compatible with that of Leibfried.

§ 2. DISCUSSION OF RAYLEIGH'S AND KING'S FORMULÆ FOR THE PRESSURE OF A SOUND WAVE.

King considers a single plane wave of wavelength λ incident on a sphere of radius a. If the energy density of the sound wave is E, the force P on the sphere is given by

$$P=\pi a^2 E L(ka), \qquad (1)$$

where

$$k=2\pi/\lambda, \qquad (2)$$

and it is easy to see that the function L must tend to the value $\frac{1}{2}$ when ka is large. King's computation leads to $L(1)=0{\cdot}425$. The distribution of scattered radiation from a large sphere may be resolved (Morse 1948) into a forward lobe which interferes with the incident beam to form the shadow and an isotropic reflected beam. Thus, of the radiation intercepted by the scattering cross-section $2\pi a^2$, half continues in the forward direction with unchanged momentum, and half is scattered isotropically with zero total momentum. If the momentum in unit volume of the incident wave is m, the change of momentum in unit time caused by the sphere is $mc\pi a^2$, and comparison with (1) leads to

$$m=E/2c. \qquad (3)$$

This appears to disagree with Rayleigh's result. Both authors assume that the pressure p in the gas depends only on its density ρ according to the law

$$p=f(\rho). \qquad (4)$$

If the density of the undisturbed gas is ρ_0, the velocity of sound c is given by

$$c^2=f'(\rho_0). \qquad (5)$$

Rayleigh's formula for the momentum is, in the present notation,

$$m=[1+\rho_0 f''(\rho_0)/2c^2]\mathrm{E}/2c. \qquad (6)$$

The discrepancy between (3) and (6) does not arise from an implied assumption $f''(\rho_0)=0$ (corresponding to Boyle's Law) in the derivation of (3), but from a difference in the physical situations considered. A similar ambiguity is known (Andrejew 1940) to arise in the determination of the energy density in a sound wave. King treats a finite obstacle in an infinite gas. Because the plane sound wave is disturbed by the obstacle, the time average of the density of the gas varies from place to place, and this variation contributes to the resultant force. Rayleigh considers a fixed quantity of air in a cylindrical pipe, either at rest or containing a resonant sound wave in a quasi-stationary state. The pipe is closed by a piston at one end and by a rigid wall at the other. Rayleigh's argument will be developed here in the notation which will be convenient in later work.

In Cartesian coordinates x_i, the components of velocity u_i are given by the equations of motion :

$$\rho\frac{\mathrm{D}u_i}{\mathrm{D}t}=-\frac{\partial p}{\partial x_i}, \qquad (7)$$

where

$$\frac{\mathrm{D}}{\mathrm{D}t}=\frac{\partial}{\partial t}+u_i\frac{\partial}{\partial x_i}. \qquad (8)$$

The repeated suffix i implies a summation over the values $i=1, 2, 3$.

If a function ϖ is introduced by the relation

$$\varpi=\int\frac{dp}{\rho}, \qquad (9)$$

the equations of motion become

$$\frac{\mathrm{D}u_i}{\mathrm{D}t}=-\frac{\partial\varpi}{\partial x_i}. \qquad (10)$$

If the motion is irrotational, there is a velocity potential ϕ satisfying

$$u_i=-\frac{\partial\phi}{\partial x_i} \qquad (11)$$

(Rayleigh here uses $-\phi$.)

The equations of motion are satisfied if

$$\varpi=\frac{\partial\phi}{\partial t}-\frac{1}{2}\left(\frac{\partial\phi}{\partial x_i}\right)^2=\frac{\mathrm{D}\phi}{\mathrm{D}t}+\frac{1}{2}\left(\frac{\partial\phi}{\partial x_i}\right)^2. \qquad (12)$$

Changing only the notation, Rayleigh's argument now proceeds: "If we integrate over a long period of time, ϕ disappears, and we see that $\int\varpi\, dt+\frac{1}{2}\int(\partial\phi/\partial x)^2\, dt$ retains a constant value at all points of the cylinder.

The value at the piston is accordingly the same as the mean value taken over the length of the cylinder." The pressure on the piston is derived from the value of ϖ at that point. According to (9), the definition of ϖ involves an arbitrary constant of integration. Rayleigh determines this constant from his "mean value" result just quoted, using his assumption that the total mass of gas, and hence the mean density taken over the length of the cylinder, is unaltered by the presence of the sound wave. The term in Rayleigh's formula for the pressure on the piston which depends only on conditions near the piston would lead to (3). The term in f'' in (6) arises from integrals over the length of the tube, expressing the constancy of the total mass of gas.

Two points in the argument quoted deserve rather more attention. Firstly, it is obvious that ϕ does not vanish in the time integral. It is necessary to take the time average before ϕ can disappear. If ϕ actually did "disappear" in the integration, the time average of $\varpi+\frac{1}{2}(\partial\phi/\partial x)^2$ would not be a determinable constant but zero, leading again to (3). In fact the arbitrary constant of integration in ϖ corresponds to an arbitrary addition to ϕ of the form $t\psi(x)$. This gives an aperiodic contribution $-t\,d\psi/dx$ to u_x, and the postulate of a steady state leads to $\psi(x)=\text{constant}$, which justifies Rayleigh's method.

The second point is that when terms of the second order in the wave equation are considered, a true stationary state almost certainly does not exist (Rayleigh 1910). Time averages can only be given a meaning by a special process. If the unperturbed motion is represented by a sufficiently small periodic function ϕ_0, the second order terms are small perturbations. After time t, ϕ becomes $\phi_0+\delta\phi$, where $\delta\phi$ is in general of order $\phi_0^2 t$. If $\delta\phi$ in fact has this form, the time average of $\partial\phi/\partial t$ does not tend to zero, but tends to an unknown quantity of order ϕ_0^2, which is of the same order as the terms to be evaluated. A periodic state may be maintained near the piston by using a train of waves in an infinite tube, but this no longer requires that the mean density should be unchanged. Alternatively, the rigid end wall of the cylinder may be replaced by another piston which is controlled from outside in an attempt to maintain a periodic state in the tube. This artifice can only be used if it can be shown to be possible, and in particular it must be shown that the piston will not continuously supply energy to the gas in the tube. The second order terms in the wave equation lead to an irreversible change, and thus entropy is continuously generated in the gas. In a steady state this entropy flows out through the walls of the tube as heat, while the controlled piston supplies energy at the same rate. This is exactly the experimental arrangement which would normally be adopted to measure the pressure of radiation in a closed tube. It is remarkable that a constant flow of mechanical work into the tube and of heat out is not a result of experimental imperfections, but is essential to a self-consistent interpretation of the results.

§3. RELATION BETWEEN THE RESULTS OF THIS PAPER AND THE KINETIC THEORY OF GASES.

Equation (1) gives the force on a sphere at rest in a single sound wave. If there is an isotropic distribution of sound waves of the same wavelength and of energy density $\bar{E}$, the force on a sphere at rest clearly vanishes. It will be shown in § 6 that if the sphere moves with velocity v it is subject to a retarding force $\bar{P}$ given by

$$\bar{P}=\pi a^2\bar{E}(v/c)\bar{L}(ka), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where

$$\bar{L}(ka)=\tfrac{1}{3}[4L(ka)+kaL'(ka)]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The first term in (14) arises from the " aberration " of the sound waves, the moving sphere meeting a larger flux of sound waves on the leading than on the trailing hemisphere. The second term arises from the Doppler effect, the waves incident from ahead having a shorter effective wavelength and a correspondingly larger scattering cross-section than those incident from behind.

Consider a large sphere moving through a monatomic gas at rest. The pressure of the gas is caused by the motion of the gas atoms. In the case of a solid it is customary to resolve the thermal motion of the atoms into normal modes of vibration of the lattice, and it is plausible to do the same in the case of a gas. In both cases it is necessary to consider vibrations of wavelengths as short as twice the interatomic spacing, and nevertheless the vibrations are regarded as those of a continuous medium. These short waves are far more numerous than the larger waves, for which such an approximation is reasonably accurate. In both cases the scattering of short waves is so large that their mean free path is of the order of one wavelength. Since this theory is reasonably successful for solids, it might be expected also to succeed for gases. If the whole thermal energy of the gas is resolved into sound waves in this way, the moving sphere is retarded according to (13), with $ka \gg 1$. If the pressure of the gas is p, then $\bar{E}=\frac{3}{2}p$, while $L(\infty)=\frac{2}{3}$. So

$$\bar{P}=\pi a^2 pv/c. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

A retardation of this enormous value would leave billiards as the only practicable ball game. While the approximations leading to (15) could alter the value of $\bar{P}$ appreciably, it is obvious that the derivation is not merely inaccurate, but basically wrong.

The same problem may be treated in a more orthodox way by regarding the pressure of the gas on the sphere as being produced, not by the impact of sound waves, but by the impact of atoms. Again, more atoms strike on the leading hemisphere than on the trailing, and, if the sphere is smooth, formula (15) is again obtained. In this case the error in the assumptions is obvious. The result can only be expected to hold if a is much less than the mean free path of atoms in the gas. The distribution of momenta among atoms striking an element of the surface of the sphere is then the superposition of two distributions, corresponding to incident and reflected streams of gas. If a is large compared with the mean free path, the distribution of momenta among atoms reaching the surface is not that corresponding to distant parts of the gas, but corresponds to the pressure and mean velocity of the gas (which is parallel to the surface) at that point in the ordinary hydrodynamical theory. Forces of the kind just considered arise only when v changes, and give rise to an increase in the effective mass of the sphere.

In the original derivation of (15), where the thermal motion of the gas was resolved into sound waves, it is now clear that the short waves, which contribute most of the pressure p, do not contribute to $\overline{\mathrm{P}}$ because of their short free path, but there remain enough long waves of very long free path to give a retardation far greater than is observed. The basic difficulty in regarding the pressure of a gas as the sum of the pressures of all the sound waves in the gas now appears. It follows from (9) and (12) that in any standing wave the time average of the pressure is greatest at the nodes, where $(\partial\phi/\partial x_i)^2=0$, and least at the antinodes. Since the wall is a node for every wave, the pressure on the wall exceeds the pressure in the interior, which is equal to the "internal pressure" of the gas, that is, zero for a perfect gas. In a solid, this resolution of the thermal motion into independent sound waves is an acceptable approximation, because each sound wave can be propagated alone through the elastic solid. This is not true in the gas. The velocity of propagation of a sound wave is proportioanl to $\mathrm{T}^{\frac{1}{2}}$, where T is the absolute temperature. If the whole thermal energy of the gas is resolved into sound waves, T is a measure of the number of sound waves already present, and the propagation of one sound wave depends essentially on the presence of others. To analyse the motion into independent sound waves is therefore meaningless. If, in the problem of a sphere moving through a gas at rest, the thermal motion of the gas is resolved into sound waves, the disturbance of one particular sound wave by the sphere is produced not only directly, but also indirectly by the interaction of this sound wave with other sound waves already disturbed by the sphere. In the hydrodynamic solution, the points of stagnation are points of high pressure and so, in adiabatic flow, of high temperature and of high sound velocity, while the equator is a region of low sound velocity. This temperature distribution refracts sound waves so that they tend to travel along the lines of flow, in the same way that, in the molecular picture, incident molecules are deviated by collisions towards the lines of flow. Thus (13) cannot be disproved by deducing from it the absurd consequence (15).

§ 4. THE BASIC EQUATIONS.

The motion of the gas is governed by the equation of continuity,

$$\frac{\partial\rho}{\partial t}+\frac{\partial(\rho u_i)}{\partial x_i}=0, \quad \ldots\ldots\ldots \quad (15)$$

which may be written

$$\frac{\mathrm{D} \ln \rho}{\mathrm{D}t} = \frac{\partial^2 \phi}{\partial x_i^2}. \quad \dots\dots\dots \quad (16)$$

It follows from (4) and (9) that ρ may be expressed as a function of ϖ alone, and then from (12) that ρ may be expressed in the form

$$\rho = \rho(\phi). \quad \dots\dots\dots\dots \quad (17)$$

The function $\rho(\phi)$ involves both ϕ and its derivatives. The equation of motion (16) may now be written in terms of ϕ alone as

$$\frac{\mathrm{D} \ln \rho(\phi)}{\mathrm{D}t} = \frac{\partial^2 \phi}{\partial x_i^2}, \quad \dots\dots\dots \quad (18)$$

while the pressure is given by

$$p = f[\rho(\phi)]. \quad \dots\dots\dots\dots \quad (19)$$

These equations will now be expanded in powers of ϕ, using as an auxiliary variable the condensation s, a pure number of the same order as ϕ, defined by

$$\rho = \rho_0(1+s). \quad \dots\dots\dots \quad (20)$$

To the third order (4) becomes

$$p = p_0 + \rho_0 c^2 s + \tfrac{1}{2}\rho_0^2 f'' s^2 + \tfrac{1}{6}\rho_0^3 f''' s^3, \quad \dots\dots \quad (21)$$

where

$$p_0 = f(\rho_0) \quad \dots\dots\dots\dots \quad (22)$$

and the derivatives f'' and f''' are evaluated at $\rho = \rho_0$. To the same order (9) becomes

$$\varpi = c^2 s - \tfrac{1}{2}(c^2 - \rho_0 f'')s^2 + \tfrac{1}{3}(c^2 - \rho_0 f'' + \tfrac{1}{2}\rho_0^2 f''')s^3. \quad \dots \quad (23)$$

Inverting (23),

$$s = \frac{\varpi}{c^2} + \tfrac{1}{2}(c^2 - \rho_0 f'')\frac{\varpi^2}{c^6} + \tfrac{1}{6}(c^4 - 4\rho_0 c^2 f'' + 3\rho_0^2 f''^2 - \rho_0^2 c^2 f''')\frac{\varpi^3}{c^{10}}. \quad . \quad (24)$$

Using (12) and (24), (17) becomes

$$\begin{aligned} \rho = \rho_0 &+ \frac{\rho_0}{c^2}\frac{\partial \phi}{\partial t} - \frac{1}{2}\frac{\rho_0}{c^2}\left(\frac{\partial \phi}{\partial x_i}\right)^2 + \frac{1}{2}\frac{\rho_0}{c^6}(c^2 - \rho_0 f'')\left(\frac{\partial \phi}{\partial t}\right)^2 \\ &- \frac{1}{2}\frac{\rho_0}{c^6}(c^2 - \rho_0 f'')\frac{\partial \phi}{\partial t}\left(\frac{\partial \phi}{\partial x!}\right)^2 \\ &+ \frac{1}{6}\frac{\rho_0}{c^{10}}(c^4 - 4\rho_0 c^2 f'' + 3\rho_0^2 f''^2 - \rho_0^2 c^2 f''')\left(\frac{\partial \phi}{\partial t}\right)^3. \quad \dots \quad (25) \end{aligned}$$

It is convenient to introduce the abbreviations

$$\frac{\partial \phi}{\partial t} = \dot{\phi}; \qquad \frac{\partial \phi}{\partial x_i} = \phi_i, \qquad \frac{\partial^2 \phi}{\partial x_i \partial t} = \dot{\phi}_i, \qquad \frac{\partial^2 \phi}{\partial x_i \partial x_j} = \phi_{ij}, \text{ etc.} \quad . \quad (26)$$

With this notation the equation of motion (18) becomes, to second order,

$$c^2\phi_{ii}-\ddot{\phi}=-2\phi_i\dot{\phi}_i-\rho_0c^{-4}f''\dot{\phi}\ddot{\phi}, \quad . \quad . \quad . \quad . \quad . \quad (27)$$

and (21) becomes, to third order,

$$p=p_0+\rho_0\dot{\phi}-\tfrac{1}{2}\rho_0\phi_i^2+\tfrac{1}{2}\rho_0c^{-2}\dot{\phi}^2-\tfrac{1}{2}\rho_0c^{-2}\dot{\phi}\phi_i^2+\tfrac{1}{6}\rho_0c^{-6}(c^2-\rho_0f'')\dot{\phi}^3. \quad (28)$$

It is noteworthy that f''', which appears explicitly in (21), is absent from (28).

§ 5. A BODY MOVING WITH UNIFORM VELOCITY THROUGH A GAS CONTAINING A SINGLE PROGRESSIVE SOUND WAVE.

In treating the interaction of a sound wave with a body moving with uniform velocity v, it is convenient to use axes $(x_1'\ x_2'\ x_3')$ travelling with the body, and also to make a Lorentz transformation of the time. The new coordinates are given by

$$\left.\begin{aligned}(1-v^2/c^2)^{\frac{1}{2}}x_1'&=x_1-vt,\\ x_2'&=x_2,\\ x_3'&=x_3,\\ (1-v^2/c^2)^{\frac{1}{2}}t'&=t-vx_1/c^2.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

If v/c is treated as a small quantity of the first order, the wave equation (27) takes the same form, to second order, in primed coordinates. In (28) the term $\rho_0\dot{\phi}$ does not contribute to the time average of p, and the remaining terms are unaltered, to third order, by the change of coordinates. (By the argument of § 2, a term in ϕ proportional to t can only lead to a change in pressure which is independent of position, and so produces no force on a closed surface.) Equations (27) and (28) may therefore be considered to hold in moving coordinates. They are "mixed" equations, in which p and ϕ retain their original values measured in fixed coordinates, while the differentiations are carried out in moving coordinates.

The present problem is concerned with the interaction between two disturbances. One, representing the uniform motion of the body with speed v, is independent of time; the other is, in first order, a sinusoidal function of frequency ω. The first order solution of (27) is then expressible in the form

$$\phi^{(1)}=\mathrm{V}(x_i)+\mathrm{C}(x_i)\cos\omega t+\mathrm{S}(x_i)\sin\omega t.$$

In moving coordinates, the disturbance represented by C and S is still periodic, and the first order solution may be written

$$\phi^{(1)}=\mathrm{V}[(x_1'+vt')(1-v^2/c^2)^{-\frac{1}{2}},\ x_2',\ x_3']+\mathrm{C}'(x_i')\cos\omega't'+\mathrm{S}'(x_i')\sin\omega't'. \quad . \quad . \quad (30)$$

If the solution to second order is $\phi=\phi^{(1)}+\phi^{(2)}$, $\phi^{(2)}$ is a solution of the equation

$$\begin{aligned}c^2\phi_{ii}^{(2)\prime}-\ddot{\phi}^{(2)\prime}=&-2\omega'(\mathrm{V}_i'+\mathrm{C}_i'\cos\omega't'+\mathrm{S}_i'\sin\omega't')(-\mathrm{C}_i'\sin\omega't'+\mathrm{S}_i'\cos\omega't')\\ &+\rho_0c^{-4}f''\omega'^3(-\mathrm{C}'\sin\omega't'+\mathrm{S}'\cos\omega't')(\mathrm{C}'\cos\omega't'+\mathrm{S}'\sin\omega't').\end{aligned}$$
$$. \quad . \quad . \quad . \quad (31)$$

Here ϕ'_{ii} is an abbreviation for $\partial^2\phi/\partial x_i'^2$, and C'_i is an abbreviation for $\partial C'/\partial x'_i$, etc. Terms in $\partial V/\partial t'$ in (31) are of the third order, and have been neglected.

All terms in $\phi^{(2)}$ are of frequency 2ω, except those arising from the term $2\omega' V'_i(C'_i \sin \omega' t' - S'_i \cos \omega' t')$, and the solution for ϕ may be written to second order as

$$\begin{aligned}\phi&=\phi^{(1)}+\phi^{(2)}\\ &=V+C' \cos \omega' t'+S' \sin \omega' t'+C^{(2)\prime} \cos 2\omega' t'+S^{(2)\prime} \sin 2\omega' t'\\ &\qquad+(VC)' \sin \omega' t'+(VS)' \cos \omega' t', \quad \ldots \quad (32)\end{aligned}$$

where $(VC)'$ is a solution of

$$c^2(VC)'_{ii}+\omega'^2(VC)'=2\omega' V'_i C'_i, \quad \ldots \quad (33)$$

and $(VS)'$ is a solution of

$$c^2(VS)'_{ii}+\omega^2(VS)'=-2\omega' V'_i S'_i. \quad \ldots \quad (34)$$

When this solution is substituted in (28), and terms of higher order than the third are neglected, it is readily seen that the time average p_m of p is given by

$$\begin{aligned}p_m=p_0&-\tfrac{1}{2}\rho_0 V_i^2-\tfrac{1}{4}\rho_0(C_i'^2+S_i'^2)+\tfrac{1}{4}\rho_0 c^{-2}\omega'^2(C'^2+S'^2)\\ &+\tfrac{1}{2}\rho_0[C'_i(VS)'_i+S'_i(VC)'_i]+\tfrac{1}{2}\rho_0 c^{-2}\omega'^2[C'(VS)'+S'(VC)']\\ &-\tfrac{1}{2}\rho_0 c^{-2}\omega'[V'_i(C'_i S-S'C)]. \quad \ldots \quad (35)\end{aligned}$$

Here p_0 is the pressure of the undisturbed gas, the second term is the pressure due to the steady flow, which gives no force and, for a symmetrical body, no couple. The next two terms give the acoustic radiation pressure, as calculated by King, and the remaining three terms give the pressure produced by the interaction of the steady flow and the sound wave.

Using equations (33) and (34), together with the fact that C and V satisfy equations of the form

$$c^2C'_{ii}+\omega'^2C=0, \quad \ldots \quad (36)$$

reduces the last three terms in (35) to the form

$$-\tfrac{1}{4}\rho_0[C'(VS)'+S'(VC)']_{ii}. \quad \ldots \quad (37)$$

It may be shown that this term gives no resultant force on the body. For a smooth function Q may be defined inside the body, which has on the surface of the body the same first and second derivatives as $[C'(VS)'+S'(VC'')]$. If $d\Sigma'_j$ is an element of area of the surface, the total force on the body in the direction x_j is

$$-\tfrac{1}{4}\rho_0 \iint Q'_{.i}\, d\Sigma'_j=-\tfrac{1}{4}\rho_0 \iiint (\partial Q'_{ii}/\partial x'_j)\, d\tau',$$

where the integral $d\tau'$ is taken through the volume of the body. This may be written

$$-\tfrac{1}{4}\rho_0\frac{\partial}{\partial x'_j}\iiint Q'_{ii}\,d\tau'=-\tfrac{1}{4}\rho_0\frac{\partial}{\partial x'_j}\iint Q'_i\,d\Sigma'_i.$$

At each point of the surface,

$$Q'_i=C'_i(VS)'+C'(VS)'_i+S'_i(VC)'+S'(VC)'_i,$$

and each of the derivatives in this expression is orthogonal to $d\Sigma'_i$ because each term C', $(VS)'$, S', $(VC)'$ represents the velocity potential of a flow with the surface of the body as a fixed boundary.

§ 6. A BODY MOVING THROUGH AN ISOTROPIC FLUX OF SOUND WAVES.

Consider a body moving with velocity v through a large space containing an isotropic flux of sound waves of frequency ω. It follows from § 5 that the third order terms in (35) do not produce a force on the body. However, the radiation is not isotropic in coordinates moving with the body, and, even in the case of a sphere or a cylinder, the pressures corresponding to the second order terms in (35) lead to a resultant force.

In a fixed coordinate system a typical sound wave has the velocity potential

$$\phi=\phi_0\cos(k_1x_1+k_2x_2+k_3x_3-\omega t), \quad . \quad . \quad . \quad . \quad . \quad (38)$$

where

$$k^2=k_1^2+k_2^2+k_3^2=\omega^2/c^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

Referred to axes moving with the body, as given by (29), the velocity potential is

$$\phi=\phi_0\cos[(k_1-\omega v/c^2)(1-v^2/c^2)^{-1/2}x'_1+k_2x'_2+k_3x'_3-(\omega-k_1v)(1-v^2/c^2)^{-1/2}t']. \quad . \quad . \quad . \quad . \quad (40)$$

Neglecting terms of order v^2/c^2, this is a wave travelling in the direction

$$\left.\begin{aligned} l'_1&=\frac{k_1}{k}+\frac{k_1^2v}{k^2c}-\frac{v}{c},\\ l'_2&=\frac{k_2}{k}+\frac{k_1k_2v}{k^2c},\\ l'_3&=\frac{k_3}{k}+\frac{k_1k_3v}{k^2c},\end{aligned}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

with frequency

$$\omega'=\omega-k_1v \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

and wave number

$$k'=\omega'/c. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Equation (1) may be written in the form

$$P=\tfrac{1}{2}\pi\rho_0\phi_0^2(ka)^2L(ka), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

and by substituting (40) in (30) and (35) it is clear that the force on a moving sphere is given by (44) with k replaced by k'. The resultant force produced by the isotropic radiation must act in the direction x_1. The force produced by the wave $(k_1k_2k_3)$ has a component in this direction of

$$l_1'\tfrac{1}{2}\pi\rho_0\phi_0^2(k'a)^2\mathrm{L}(k'a). \qquad (45)$$

On setting $l_1=k_1/k$, this may be written, to first order in v/c, as

$$\tfrac{1}{2}\pi\rho_0\phi_0^2\{l_1k^2a^2\mathrm{L}-(v/c)[(1+l_1^2)k^2a^2\mathrm{L}+l_1^2k^3a^3\mathrm{L}']\}. \qquad (46)$$

If $\phi_0\, d\Omega$ is now taken to represent the velocity potential of waves in a small solid angle $d\Omega$, the total force $\overline{\mathrm{P}}$ on the sphere is obtained by integrating (46) over all directions as

$$\overline{\mathrm{P}}=-\frac{2\pi^2\rho_0\phi_0^2k^2a^2v}{3c}[4\mathrm{L}(ka)+ka\mathrm{L}'(ka)]. \qquad (47)$$

The energy density in the sound flux is

$$\overline{\mathrm{E}}=2\pi\rho_0\phi_0^2k^2, \qquad (48)$$

and (47) may thus be written

$$\overline{\mathrm{P}}=-\frac{\pi a^2\overline{\mathrm{E}}v}{3c}[4\mathrm{L}(ka)+ka\mathrm{L}'(ka)]. \qquad (49)$$

This agrees with (13) and (14).

REFERENCES.

ANDREJEW, N., 1940, *J. Phys. U.S.S.R.*, **2,** 305.
FRANK, F. C., 1948, 1947 *Bristol Conference Report*, p. 46 (London: The Physical Society).
KING, L. V., 1934, *Proc. Roy. Soc.* A, **147,** 212.
LAMB, SIR H., 1932, *Hydrodynamics* (Cambridge).
LEIBFRIED, G., 1950, *Z. Physik*, **127, 344.**
MORSE, P. M., 1948, *Vibration and Sound*, p. 355 (New York : McGraw-Hill)
RAYLEIGH, Lord, 1905. *Phil. Mag.*, [vi], **10,** 364 ; 1910, *Ibid.*, [vi], **20,** 450.

CXIV. *A Note on the Decay of Current in Germanium Diodes.*

By J. R. Tillman and H. Yemm,
The Post Office Engineering Research Station, London*.

[Received October 4, 1950.]

A modulator using germanium crystal diodes, for use in time division multiplex systems, shows an effect which enables a property of the crystal diode—not hitherto reported on—to be investigated.

Fig. 1 (*a*) shows part of a multiplex system consisting of two channels only. $(E_i)_1 = A \sin \omega_1 t$ is the source of modulation for channel 1. At all times other than when a P1 pulse is applied the diode W1A passes about 1 ma. and has low impedance, R_f ($\simeq 150\Omega$), to a.c.; the diode W1B is therefore subject to a " reverse " voltage of about 1 and presents a high impedance, R_r, to a.c. When a P1 pulse is applied W1B passes a pulse of current whose value lies between 2·2 and 3·8 ma. according to the instantaneous value of $A \sin \omega_1 t$; W1A becomes of high impedance. The modulator of channel 2 works in a similar way, with $\sin \omega_2 t$ for $\sin \omega_1 t$, P2 for P1, etc.; its output is connected in parallel with that of channel 1. The pulse length is fixed at 0·2 μsec. but the relative timing of P1 and P2, t_c, is controllable. The trains of pulses, designated HP1, HP2, are together transmitted to the receiving end of the system where they are separated and demodulated by low pass filters. The two outputs $(E_0)_1$ and $(E_0)_2$ substantially reproduce the waveforms of $(E_i)_1$ and $(E_i)_2$ respectively, with equal efficiency. Each, however, contains components, at low levels, in addition to that wanted. These include a crosstalk component; thus $(E_0)_1$ contains a component C_{21} in $\omega_2 t$ and $(E_0)_2$ a component C_{12} in $\omega_1 t$.

The circuit is so designed that the crosstalk ratio C_{12}/E_0, expected, is small; the crystal diode W1B is chosen to have $R_r > 1\text{M}\Omega$ in order that the component of C_{12} arising from the finite values of R_f and R_r is $< E_0/5000$. Only for values of $t_c < 0{\cdot}5$ μsec. will the attenuation distortion at high frequencies, introduced by residual shunt capacitances in the common path, cause C_{12} to exceed $E_0/5000$.

However, in practice, values of C_{12} of up to $E_0/100$ have been noted at $t_c = 1{\cdot}0$ μsec., depending on the sample of diode used as W1B. These values decay approximately exponentially, falling to 1/eth in about 0·4 to 1·2 μsec. (see fig. 2). The use of either a thermionic diode or some other samples of crystal diodes as W1B reduces the crosstalk to about the expected low level residual.

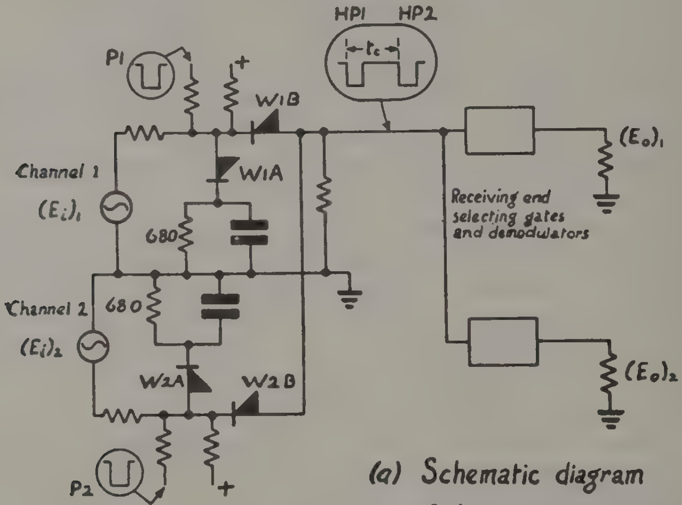
The newly noted crosstalk voltage is always out of phase, by π, with the signal obtained with $t_c = 0$, and hence with any leak due to the finite value of R_r. Any failure on the part of R_r rapidly to reach its steady state value after the cessation of P1 is, therefore, not the cause of the new

---

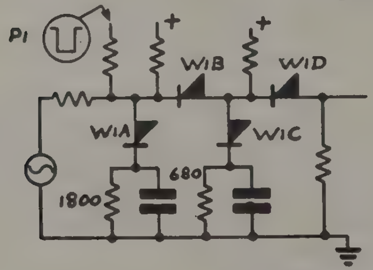
* Communicated by the Author.

crosstalk voltage. This conclusion is supported by experiments with the modulator of channel 1 changed to that of fig. 1 (*b*). Here W1B is a thermionic diode, or a crystal giving no significant crosstalk when used as W1B in the circuit of fig. 1 (*a*). Crosstalk still arises when a crystal giving crosstalk in the position of W1B of fig. 1 (*a*) is used as W1D in fig. 1 (*b*), despite the fact that signal leak is now adequately suppressed by W1A and W1B.

Fig. 1.


HP1 HP2
t_c
P1
+
W1B
W1A
Channel 1
$(E_i)_1$
680
Channel 2
680
$(E_i)_2$
W2A
W2B
P2
+
$(E_o)_1$
Receiving end selecting gates and demodulators
$(E_o)_2$

(*a*) Schematic diagram of the multiplex system


P1
+
+
W1B
W1D
W1A
W1C
680
1800

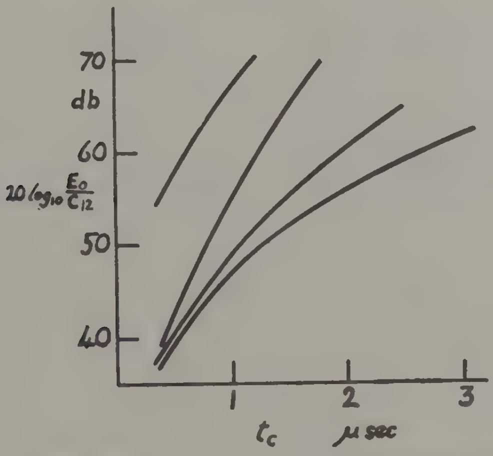
(*b*) The modified modulator

The explanation of the new crosstalk seems to lie in the slowness of dispersal of the charge carriers. But the evidence available on transit times does not afford a simple explanation.

Lark Horovitz (1946) has noted that the rise of current through a germanium diode is noticeable slower than the rise of an applied voltage, when the latter takes place in 10^{-7} sec. ; the decay of his voltage pulses was so slow as to obscure the effect now noted. Douglas and James (in press) suggest that the finite transit time of the charge carriers may be responsible for the decrease of rectification efficiency noted at frequencies above about 10 Mc/s. They have correlated this falling-off with increasing "turnover" voltage, V_t, and have pointed out that turnover voltage may well be a measure of the thickness of the barrier layer. They do not give a value to any transit time, but in view of the fact that the rectification efficiency is not grossly impaired at 50 Mc/s., a time of 10^{-8} sec. rather than 10^{-6} sec. seems to be indicated.

Correlation between the crosstalk voltage and other properties of the crystal diodes was sought, even though the large differences of crosstalk voltage noted (see fig. 2) could be obtained from units differing little in V_t. In general, the crosstalk voltage was smaller the lower the turnover voltage; two batches of crystal rectifiers type CG5 supplied by Mr. Kinman of the B.T.H. Co., Rugby, having V_t less than 80 volts, consistently gave $C_{12}<E_0/2000$ at $t_c=1\,\mu$sec., whereas some CG4's ($V_t>100$ volts) gave $C_{12}=E_0/200$ at $t_c=1\,\mu$sec. At the same time some other CG4's gave $C_{12}<E_0/3000$. Units with low turnover voltages often have low values of R_r. There seemed no correlation between C_{12} and R_r, however; in fact, one of the two batches of CG5's giving small values of C_{12} had, in the main, values of R_r of about 100kΩ, whereas the other batch was specially made to have values of $R_r>1$ MΩ.

Fig. 2.



The variation of C_{12} with time for four samples of diode.

Some of these experiments point, therefore, to the effect being independent of the operation of the diodes which gives them their better known properties; further work based on this indication is in hand.

REFERENCES.

HOROVITZ, K. LARK, 1946, National Defence Research Committee Report 14–585, "Preparation of Semi-Conductors and Development of Crystal Rectifiers."

DOUGLAS, R. W., and JAMES, E. G., *Proc. Inst. Elect. Eng.* To be published shortly.

CXV. CORRESPONDENCE.

Superconductivity of Niobium.

By L. C. Jackson and H. Preston-Thomas*.
H. H. Wills Physical Laboratory, University of Bristol.

[Received October 10, 1950.]

Until recently, published data on the properties of the "hard" superconductors were both scanty and inconsistent, particularly in the case of niobium. Studies have been made in this laboratory of the properties of the latter metal, both before and after heat treatment. At the start of the investigation the only published measurements of the critical magnetic field of niobium were those of Daunt and Mendelssohn (1937) in the liquid helium region. A cryostat was therefore constructed to extend the measurements to the region between 4·2° K. and the normal transition temperature, given by Meissner, Franz and Westerhoff (1933) as 9·2° K. A description of this apparatus is to appear elsewhere. Three methods have been used to determine the superconducting properties; resistance measurements, an A.C. inductance method identical in essentials with that of Webber, Reynolds and McGuire (1949) and the ballistic method of Keeley and Mendelssohn (1936). The satisfactory operation of the apparatus was checked by carrying out measurements on pure lead (H.S. Lab. no. 11,116) by the three methods mentioned above.

The critical magnetic fields obtained by the ballistic method for a sample of niobium (Johnson, Matthey and Co. H.S. Lab. no. 1673—spectroscopic examination showed the presence of traces of molybdenum, nickel and silicon together with a very slight trace of aluminium—used as received, not heat treated) are given in fig. 1; curve A, beginning of penetration of magnetic field, curve B completion of penetration. The amount of frozen-in flux is not large, of the order 10 to 15 per cent and varies with the temperature. Curve C of fig. 1 is derived from the alternating magnetic field measurements and agrees closely with the curve determined from measurements of the electrical resistance of niobium tape rolled from the same sample. The spread of the transition curve in the absence of a magnetic field is about 0·07 deg.

---

* Communicated by the Authors.

Curves A′, B′ and C′ of fig. 1 give the corresponding results for another sample of niobium. (Johnson, Matthey and Co. H.S. Lab. no. 2088—trace of silicon and very slight trace of aluminium detected spectroscopically—used as received, no heat treatment.) Curve D, fig. 2 is the critical field curve for niobium recently determined by Cook, Zemansky and Boorse (1950) using the alternating magnetic field method.

Fig. 1.

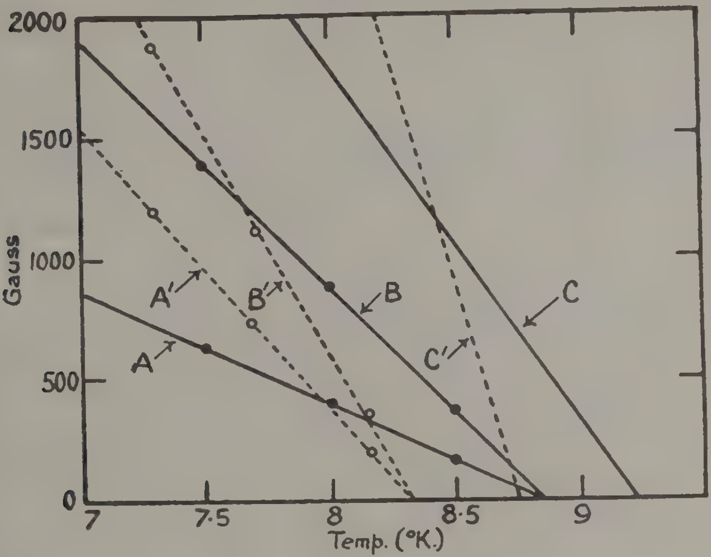


Fig. 2.

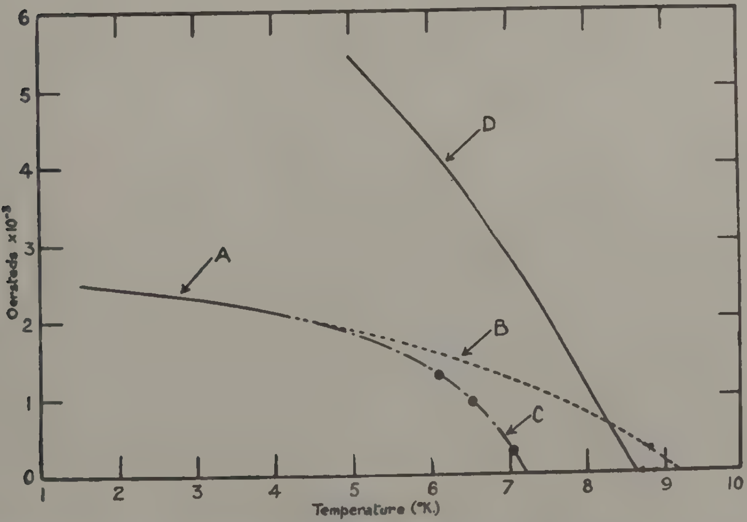


Fig. 2 shows the results of the ballistic method for an older, less pure specimen of niobium (Hilger H.S. Lab. no. 8290—spectroscopic analysis showed the presence of about 0·2 per cent tin and possibly slight traces of copper, iron and gold—used as received, not heat treated). This is also the Lab. no. of the sample used by Daunt and Mendelssohn (1937) in their determination of the critical magnetic field between 1·5 and 4·2° K. Their measurements showed that about 20 per cent of the critical field remained frozen in the specimen. The solid line A in fig. 2 gives their results. The dotted line B is the interpolation between 4·2° K, and the normal transition temperature determined by Meissner *et al.* (1933) as given in Shoenberg's "Superconductivity" (1938). The points on the dotted line C are those at which the penetration of the magnetic field into the specimen is complete. Alternating magnetic field measurements with the same specimen gave a normal transition beginning at approximately 8° K. and becoming complete at approximately 7° K. The spread of the transition appears to become smaller as the applied field is increased but it is still too great to provide useful data for the H_c—T curve. The material is, moreover, so brittle that it is impossible to roll tape specimens from it suitable for resistance measurements.

There is a considerable similarity between the above results for niobium and recently published observations on vanadium by the alternating magnetic field method (Webber *et al.* 1949) and the ballistic method (Wexler and Corak 1950).

The form of the results given above—for samples 1673 and 2088—can be explained on the basis of an extension of Mendelssohn's model of an "alloy type" superconductor (Mendelssohn 1935). This considers such a superconductor to have filaments running through it which have a much higher critical field (H_f) than that characteristic of the bulk of the material (H_c). This results in the internal field of the specimen lagging behind that which would occur in the absence of the filaments by an amount (H_a) depending on the capacity of the filaments to carry resistanceless currents.

If we now also postulate a higher normal transition temperature (T_{cf}) for the filaments than that for the bulk metal (T_c) and an inhomogeneous distribution of the frozen-in flux the following results are obtained.

The value of H_a is a function of the temperature. A graph of the frozen-in flux, which in our use of the ballistic method is equal to $2H_a$, against temperature shows that for the two niobium specimens (nos. 1673 and 2088) for which complete data have been obtained this function may be represented by $H_a=K(T_{cf}-T)$ where K is a constant. In alternating magnetic field measurements the width of the transition curve is a function of dH_a/dT.

The critical field of the filament material is given by the alternating magnetic field (and resistance) curve (curve C) and that of the bulk material by the higher of the ballistic curves (curve B). The transition temperatures and the initial slopes of the critical field curves are given in Table I.

TABLE I.

Sample	T_c°K.	T_{cf}°K.	$\frac{dH_f}{dT}$	$\frac{dH_c}{dT}$	$\frac{dH_a}{dT}$	n
1673	8·85	9·20	1,300	1,050	45	0·55
2088	8·35	8·75	4,000	1,900	60	0·39

The inhomogeneous distribution of field in the material results in an effective bulk demagnetizing factor ($4\pi n$) for the specimen. If the higher and lower of the two values of the critical field slopes given by the ballistic measurements (curve B and curve A) are dH_c/dT and dH_b/dT respectively we have $dH_b/dT=(1-n)(dH_c/dT)$.

REFERENCES.

COOK, D. B., ZEMANSKY, M. W., and BOORSE, H. A., 1950, *Phys. Rev.*, **78,** 820.

DAUNT, J. G., and MENDELSSOHN, K., 1937, *Proc. Roy. Soc.* A, **160,** 127.

MEISSNER, W., FRANZ, H., and WESTERHOFF, H., 1932, *Ann. der. Phys.*, **17,** 593.

MENDELSSOHN, K., 1935, *Proc. Roy. Soc.* A, **152,** 34.

WEBBER, R. T., REYNOLDS, J. M., and MCGUIRE, T. R., 1949, *Phys. Rev.*, **76,** 293.

WEXLER, A., and CORAK, W. S., 1950, *Phys. Rev.*, **79,** 737.

Lattice Sums for Ionic Crystals.

By F. C. FRANK,

The H. H. Wills Physical Laboratory, University of Bristol.*

[Received November 1, 1950.]

THE most important of lattice sums have by now been evaluated, at the cost of considerable labour: but there are still many which one often needs or would like to know, especially for bounded lattices or for off-lattice points. It is not too late to welcome an improved method.

In so far as the ionic charges can be regarded as distributed with spherical symmetry about the lattice points, they may be replaced (so far as the electrostatic field at any exterior point is concerned) by any other spherical distributions about the same points, or by point charges. This note is concerned with the evaluation of the potentials of the resulting lattice arrays of positive and negative point charges.

Consider the NaCl lattice as an example. Let space be divided into cubical cells, each having one ion at its centre. The point charge (positive, say) is equivalent to the sum of two charge groups: (*a*) 8 point charges of $e/8$, placed at the corners of the cell, and (*b*) one charge of e

---

* Communicated by the Author.

at the centre together with eight of $-e/8$ at the corners. We require the sum of the lattice sums for charge systems (*a*) and (*b*). The summing of (*a*) is easy, since the charges cancel at all interior points and in all cube faces and cube edges of the crystal. In a cuboid bounded by $\{001\}$ faces there remain from (*a*) only eight charges of $\pm e/8$ at the corners, situated $\frac{1}{2}\sqrt{3}\,.\,d$ cube-diagonally outwards from the corner lattice-points, and having the signs of the corner ions. An [0, 1] step in the cube face has no charge, but a Kossel kink in the step has four charges of $\pm e/8$, situated at the corners of a square of side d (the interionic distance) midway between the last ion in the incomplete step and the first missing one ; *i. e.* at the corners of the end face of the cube cell surrounding the last ion. The sign of these charges is that of the last ion.

The charge groups (*b*) are equivalent to multipole systems at the lattice points, those of lowest order being hexadecapoles. Their potentials therefore decay with distance like r^{-5}. Moreover they are of alternating sign. The lattice sum is therefore highly convergent, and can be neglected save for a few lattice points close to the point of observation. However, it is not necessary to evaluate these multipole lattice sums as such. One may cut out a cuboid containing the non-negligible lattice points and evaluate by direct summation the potential due to the ions within it, together with charges of $\pm e/8$ at its corners (each opposite in sign to the ion in that corner), and at any other corners and kinks of the crystal. This is the required sum. If the calculation is to be physically realistic there are of course surface double layers, due to the Verwey effect, and additional multipoles due to ionic distortions which also have to be brought into account. These effects apart, the corner and kink charges arising from (*a*) reproduce exactly the charge, dipole, quadrupole and octopole moments of a bounded crystal.

To test the method we use it to evaluate

$$\log_e 2 = 0{\cdot}693\ 147\ 181\ \ldots .$$

This corresponds to the potential at the next lattice point of a semi-infinite one-dimensional sodium chloride crystal. By our method the $(n+1)$th approximation (requiring the addition of $n+1$ terms) is

$$\sum_{p=1}^{n} (-1)^{p+1} p^{-1} + [(2n+1)^2+2]^{-\frac{1}{2}},$$

$n=0$ (no summation) gives 0·57735, $n=1$ gives 0·69849, $n=4$ gives 0·693085, $n=8$ gives 0·693145 and the 6th place would then be correctly estimated by inspecting the alteration of terms.

As a second test we evaluate Madelung's sum for NaCl, which gives the potential at a lattice point due to the rest of the ions in an infinite crystal. Madelung evaluated this as 1·74(2) and Emersleben as 1·74755(7±2). Treating no ions individually we have

$$2/\sqrt{3} = 1{\cdot}1547\ \ldots$$

Treating individually the first cube shell of ions we have

$$-6+12/\sqrt{2}-8/\sqrt{3}+2/3\sqrt{3}=1{\cdot}748619\ldots$$

Taking one more shell (a total of 10 terms) gives 1·747626. The four term sum is sufficiently accurate for most physical applications.

Besides possessing the merits of simple physical interpretation and easy application to bounded crystals or to summation for off-lattice points, the method appears to be more efficient for computation than any hitherto suggested, whether of higher or lower mathematics. The former methods provide admirably convergent series, but the labour comes in interpolating values of uncommon functions which are not tabulated at close intervals. The neutral shell summation methods of Evjen and Slater (split shells) and of Højendahl (eccentric shells) converge much more slowly, having errors of nearly 2 per cent (1·77681) and 0·3 per cent (1·75177) after adding 9 and 12 terms respectively.

When this method is applied to lattices of lower symmetry, the residual charge groups (*b*) will have multipoles of lower order, *e. g.* quadrupoles, making the convergence of the method slower (alternating r^{-3} terms). It may then be profitable to repeat the device, so placing quadrupoles at kink and corner points to secure rapid convergence in the residue.

REFERENCES.

EMERSLEBEN, O., 1923, *Phys. Zeits.*, **24,** 73, 79.
EVJEN, H. M., 1932, *Phys. Rev.*, **39,** 680.
HØJENDAHL, K., 1938, *K. Danske. Videns. Selskab.*, **16** (2), 138.
MADELUNG, E., 1918, *Phys. Zeits.*, **19,** 524.
PAULING, L., 1945, *Nature of the Chemical Bond* (Ithaca N.Y.: Cornell Univ. Press), p. 338 (contains general references).
SLATER, J. C., 1939, *Introduction to Chemical Physics* (New York: McGraw-Hill), 387.
VERWEY, E. J. W., 1946, *Rec. Trav. Chim. des Pays-Pas*, **65,** 521.

October 7, 1950.

Some Extensions in the Use of Resistance Thermometry in the Study of Gaseous Explosions.

By A. S. LEAH, C. ROUNTHWAITE and D. BRADLEY,
University of Leeds*.

[Received October 20, 1950.]

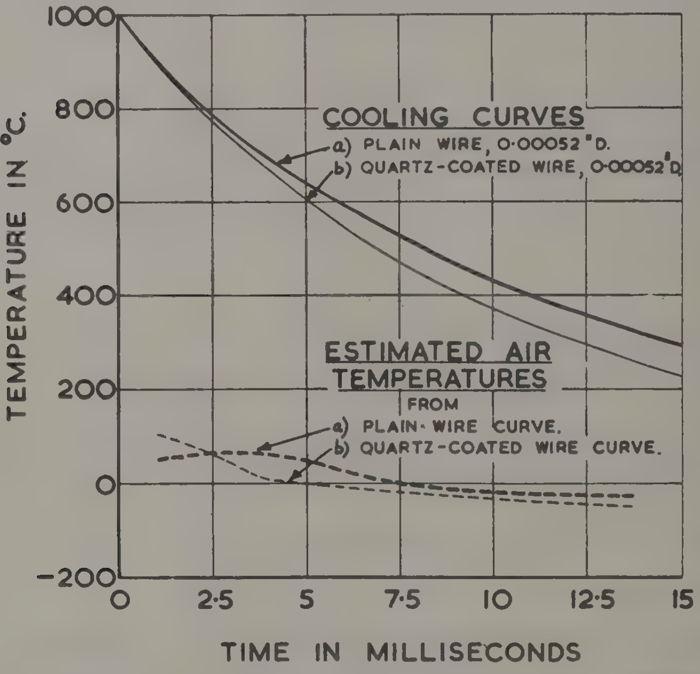
IN a recent communication to the *Philosophical Magazine* (1950, **41**, 729–30) Messrs. Daish, Fender and Woodall have raised some interesting points concerning a method we described in a recent paper to this Journal (1950, **41**, 468–77) for determining flame gas temperatures by the resistance thermometer method. For reasons of economy in space we were constrained to omit a lengthy discussion of certain aspects of the work, but it would now seem profitable to add a few words in amplification.

---

* Communicated by the Authors.

We are entirely in agreement with them when they assert that there will be considerable distortion of the temperature–time curves due to lag in response of the wires. Indeed, it was for this reason that full allowance was made for the effect as they will find by referring to expression (1) on page 472 of our paper. In actual fact, the thermal capacities of the two types of wires are very little different and although the effect of the quartz coating might be expected to increase the time constant of the coated wires, owing to the low conductivity of quartz, this is hardly an appreciable effect as we found from the results of a simple experiment.

Fig. 1.



A quartz-coated wire and then a plain platinum wire, each of 0.0005 inch diameter, were heated electrically up to a high temperature in stagnant air at atmospheric pressure within a closed vessel. A rapid-acting relay switch was made to cut off the heating current and switch the wire into the oscillograph recording circuit described in the previous paper. The oscillograph trace gave the cooling curves in each case and these, with the temperature scale, are reproduced in fig. 1. It will be noted that the cooling rates are not very different, in fact the time constants are 0.0100 sec. for the coated wire and 0.0115 sec. for the plain wire.

These tests, however, provide a very excellent check upon the accuracy of our method of correction, as will be seen from the broken line curves of fig. 1, which give the temperature of the surrounding air as deduced by exactly the same analysis as that employed in the estimation of flame gas temperatures. This temperature should have been the constant value of approximately 20° C. The discrepancy is not large even under the exacting conditions at the beginning of our records when the wire temperature was of the order of 1000° C. In fact, these experiments gave us full confidence in the method of correction we used.

Thus, we disagree with the speculation of Messrs. Daish, Fender and Woodall that there is a large difference in the time constants of the two wires and that this difference could account for the catalytic heating we observed on the plain platinum wires, especially as full correction for the small difference there is in time constant was made.

Concerning the effect of cooling of the wires by the heavier leads there is little we need add to what has already been said. We chose a length of 1 inch for our wires not for ease of measurement but because repeated tests had shown that with lengths of 1 inch or more, the cooling effect was negligibly small for the quartz-coated wires. For plain wires, calculations and tests have shown that the mean temperature exceeds 97 per cent of the maximum temperature, suggesting an error of 3 per cent. But in our experiments this error is automatically corrected because the method of calibration simulates the explosive conditions very closely in so far as cooling by the leads is concerned (the leads, of course, do not heat up appreciably in 2 seconds in view of the large time constant).

The problem of determining the true convection losses from wires in the explosions is one with which we have experienced difficulty. It is true that in the explosion there will be a certain degree of forced convection past the wires, as we pointed out previously. In addition, it may be that the equilibrium natural convection conditions have not had time to become fully established, although we have photographic evidence of the bodily convective rise of the whole flame volume in times of only a few milliseconds. Nevertheless, experiments made by Mr. N. Carpenter in these laboratories (unpublished) have shown that the heat loss from electrically heated, 0.0005 inch diameter wires, in a stream of flame gas (at varying temperatures) hardly varies at all with gas velocity past the wires up to values of 30 in. per sec. This phenomenon appears to indicate that the loss of heat is mainly dependent upon conduction through a stagnant film of gas around the wires. In view of this, we feel confident that our assumptions regarding the convective heat loss will involve no substantial error in the derived temperatures and catalytic heating rates.

CXVI. *Notices of New Books and Periodicals received.*

Structure of Molecules and the Chemical Bond. By J. K. Syrkin and M. E. Dyatkina, translated and revised by M. A. Partridge and D. O. Jordan. (Butterworths Scientific Publications.)

This book describes the properties of the chemical bond and interprets them in terms of quantum mechanics.

It is clearly addressed mainly to experimental workers and therefore starts with an account of the quantum mechanics necessary to interpret the experimental results, the less essential part of the mathematics being relegated to an appendix. There are two different approximate solutions of the wave equation that will give a semi-quantitative interpretation of the chemical bond; these are the electron-pair bond plus resonance approximation, and the molecular orbital approximation. Both these approximations are described and compared, but in the subsequent discussion of experimental results it is usually the first approximation that is used. The latter chapters of the book consist of a large amount of experimental data on the chemical bond and the interpretation of this data in terms of the theory given in the first chapters. The subjects considered include bond lengths, bond energies, dipole moments, van der Waals forces, and the general arrangement of bonds in complex molecules and crystals. A minor criticism is that in Chapter 10 the word polarization and the symbol P are used in an unusual sense that makes the discussion rather puzzling at first.

Although this book is addressed mainly to experimentalists, the theoretical worker will also find the collection of experimental data of great value. A. F. D.

Calculating Instruments and Machines. By D. R. Hartree. (Urbana: University of Illinois Press.)

The English edition of this book is now obtainable, price 21*s*. Published by Cambridge University Press, London.

BOOKS RECEIVED.

A new scientific publication has been received—‘ Geochimica et Cosmochimica Acta ’, edited with the assistance of an Advisory Board. Published by Butterworth-Springer Ltd. (London). Price per vol. (approx. 300 pp.) £3 10*s*. 0*d*.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

INDEX TO VOL. XLI.

ADAMS (N.), temporary increase in neutron compound of cosmic rays, 503.

——, —— (with Braddick, H. J. J.), 505.

Atkins (K. R.) and Osborne (D. V.), the velocity of second sound below 1° K., 1078.

Alpha particles, short range emission.—II. and III. (Burcham and Freeman), 337, 873.

——, ——.—IV. (Freeman), 1225.

——, ——, from break up of ^{8}Li (Baxter, Burcham and Paul), 937.

Alternant hydrocarbons, molecular-orbital calculations (de Heer), 370.

Atmospheric absorption of heat radiation by water vapour (Cowling), 109.

Atmospherics, use to study very long radio waves (Gardner), 1259.

Atomic oxygen in CO explosions (Leah, Rounthwaite and Bradley), 478.

Avery (D. G.), some optical properties of evaporated layers of silver, copper and tin, 1018.

Baxter (A. S.), Burcham (W. E.) and Paul (E. B.), short range alpha particles from the break up of ^{8}Li, 937.

Benson (A. C.), *see* Runcon *et al.*, 783.

Berz (F.), potentials derived from axial values in electron optics, 209.

Betatron-starting in electron synchrotrons (Wilkins), 34.

Books, new:—Bangham (D. H.), Progress in coal science, 1197; Born (M.) and Green (H. S.), General theory of liquids, 506; Bridgman (P. W.), The nature of physical theory, 619; Butterfield (H.), Origins of modern science, 107; Cauchois (Yvette), Les spectres de rayons X et la structure electronique de la matière, 731; Centre National de la Recherche Scientifique colloquia, 730; Chandrasekhar (S.), Radiactive transfer, 855; Computation Laboratory, Tables of inverse hyperbolic functions, 620; Crowther (J. A.), Ions, electrons and ionizing radiations, 208; Dingle (H.), Practical applications of spectrum analysis, 1196; Dyson (G.), Organic chemistry for advanced students, 856; Fishenden (M.) and Saunders (O. A.), Introduction to heat transfer, 732; Geochimica et Cosmochimica Acta, 1292; Hartree (D. R.), Calculating instruments and machines, 1292; Haward (R. N.), Strength of plastics and glass, 312; Jánossy (L.), Cosmic rays, 856; Johnson (M.), Astronomy of stellar energy and decay, 1197; Jessop (H. T.) and Harris (F. C.), Photoelasticity, 620; (Kniper (G. P.), Atmospheres of the earth and planets, 508; Kruyt (H. R.) Colloid science, 312; Lanczos (C.), Variational principles of mechanics, 620; Von Lane (M.), History of Physics, 856; Lothian (G. F.), Absorption spectrophotometry, 855; Martin (T.), Faraday's discovery of electromagnetic induction, 208; Milne (W. E.), Numerical calculus, 619; Milne-Thomson (L. M.), Theoretical hydrodynamics, 411; M.I.T. Technical report, supersonic flow around cones, 107; Mott (N. F.) and Massey (H. S. W.), Theory of atomic collisions, 108; Partington (J. R.), Advanced treatise on physical chemistry, Vol. I., 107; Reiner (M.), Twelve lectures on theoretical rheology, 412; Richardson (E. G.), Dynamics of real fluids, 732; Rides (J. F.), Electronic Engineering Master Index 1947–8, 619; Schelkunoff (S. A.), Applied mathematics for engineers and scientists, 108; Shannon (C. E.) and Weaver (W.), Mathematical theory of communications, 508; Sidgwick (N. V.), The chemical elements and their compounds, 731; Stoker (J. J.), Non linear vibrations in mechanical electrical systems, 731; Syrkin (J. K.) and Dyatkina (M. E.), Structure of molecules and the chemical bond, 1292; Tolansky (S.), Introduction to atomic physics, 208; Treloar (L. R. G.), Physics of rubber elasticity, 412; Wiener (N.), Stationary Time Series with engineering applications, 411.

Bourne (L.) and Hill (R.), rolled sheet in tension and cupping tests, 671.

Bowden (K. F.), effect of eddy viscosity on ocean waves, 907.

Braddick (H. J. J.), *see* Adams and Braddick, 505.

Bradley (D.), *see* Leah *et al.*, 468, 478, 1289.

Brinkley (T. A.), *see* Titterton and Brinkley, 500.

Buchdahl (H. A.), thermal equilibrium in gravitational sphere fluid, Tolman's equation, of 362.

Burcham (W. E.) and Freeman (Joan M.), emission of short range alpha particles from light elements under proton bombardment.—II., 337.

——, —— .—III., 873.

——, *see* Baxter, Burcham and Paul, 937.

Burge (E. J.) and Jackson (L. C.), transfer rates for surface films of HeII, 205.

Burton (P. C.), *see* Farnell, Burton and Hallama, 157, 545.

Calnan (E. A.) and Clews (C. J. B.), deformation textures in face centred cubic metals, 1085.

Camerini (U.), Fowler (P. H.), Lock (W. O.) and Muirhead (H.), nuclear transmutations produced by cosmic particles of great energy.—Part IV., 413.

Canetti (G. S.), *see* Whitehead and Canetti, 988.

Carlson (A. G.), Hooper (J. E.) and King (D. T.), cosmic ray particles of great energy.—Part V. The neutral mesons, 701.

Chess, programming a computer for playing (Shannon), 256.

Chester (W.), propagation of sound waves in an open-ended channel, 11.

Clegg (J. A.) and Davidson (I. A.), radio echo method for measurement of heights of meteor trails, 77.

Clews (C. J. B.), *see* Calnan and Clews, 1085.

Cockroft (A. L.), *see* Curran *et al.*, 517.

Coleman (B. L.), propagation of E.M. disturbances along a thin wire in stratified medium, 276.

Communication theory and physics (Gabor), 1161.

Compressible fluid, velocities and local Mach numbers (Tempest), 382.

Conn (G. K. T.), *see* Donovan and Conn, 770.

Corinaldesi (E.) and Field (G.), scattering of pseudoscalar charged mesons by nucleons.—II., 364.

Cosmic Radiation, heavy nuclei (Dainton and Kent), 963.

—— ——, mass of charged particles in (Franzinetti), 86.

—— ——, nuclear transmutations by particles of great energy.—Pt. III. (Fowler), 169.

—— —— .—Pt. IV. (Camerini, Fowler *et al.*), 413.

—— —— .—Pt. V. (Carlson, Hooper and King), 701.

—— ——, primary, cut off in energy spectrum (Gething), 495.

—— ——, stars in emulsions, gap measurement (Hodgson), 725.

—— ——, protons and α particles in stars (Perkins), 138.

Cosmic rays, temporary increase in neutron component (Adams), 503; (Adams and Braddick), 505.
—— ——, double stars (Li), 1152.
—— ——, stars, latitude effect (Lattimore), 961.
—— ——, single protons, origin (Lattimore), 819.
—— ——, momentum and sign of fast particles (Moral), 1058.
Coster (D.) and Kiestra (S.), fine structure near the Fe K-absorption edges in various oxides of Fe, 144.
Cowling (T. G.), atmospheric absorption of heat radiation by water vapour, 109.
Curran (S. C.), Cockroft (A. L.) and Insch (G. M.), investigation of soft radiation.—VII. Energy per ion-pair for slow electrons in gases, 517.
Dainton (A. D.) and Kent (D. W.), heavy nuclei of cosmic radiation, 963.
Daish (C. B.), Fender (D. H.) and Woodall (A. J.), use of resistance thermometers for rapidly changing temperatures, 729.
Daniel (Vera), order disorder transitions in mixed ketones, 631.
Darwin (Sir C.), some conformal transformations involving elliptic functions, 1.
Davidson (I. A.), *see* Clegg and Davidson, 77.
Davies (I. L.), *see* Woodward and Davies, 1001.
Deformation, inhomogeneous, of a plastic lamina (Hill), 733.
——, textures in metals (Calnan and Clews), 1085.
Density variations in aluminium (Snoek), 1188.
Di-iso-electronic row (Richter), 983.
Dilworth (Miss C. C.), Goldsack (S. J.), Goldschmidt-Clermont (Y.) and Levy (F.), magnetic deflection of fast charged particles in the photographic emulsion, 1032.
Domb (C.), the statistics of correlated events.—I., 969.
Donovan (B.) and Conn (G. K. T.), electrical conductivity of bismuth fibres.—II. Anomalies in the magneto-resistance, 770.
Eddy viscosity, in turbulent shear flow (Townsend), 890.
—— ——, effect on ocean waves (Bowden, 907.
Elasticity, theory in general coordinates (Green and Zerna), 313.
Electrical conductivity of bismuth fibres.—II. (Donovan and Conn), 770.
—— noise, statistical analysis (MacDonald), 814.
Electromagnetic disturbances, propagation along thin wire (Coleman), 276.
Electron optics, potentials derived from axial values (Berz), 209.
—— streams, transit-time.—II. (MacDonald), 863.
Elliptic functions, conformal transformations involving (Darwin), 1.
Ellyett (C. D.), influence of high altitude winds in meteor trail ionization, 694.
Energy distributions behind decaying shocks.—I. (Lighthill), 1101.
—— release in reaction ^{7}Li$(dp)^8$Li (Paul), 942.
—— states, calculation by uncertainty relations (Soh), 851.
Evaporated layers, optical properties (Avery), 1018.
——, interferometric study (Faust), 1238.
Farnell (G. C.), Burton (P. C.) and Hallama (R.), fluorescence of silver halides at low temperatures.—Pt. I. The pure halides, 147.
——, —— .—Pt. II. Mixed halides, 545.
Faust (R. C.), interferometric study of evaporated silver films, 1238.
Fender (D. H.), *see* Daish *et al.*, 729.
Ferromagnetics, domain boundaries (Lilley), 792.
Field (G.), *see* Corinaldesi and Field, 364.
Fine structure near K absorption edges in oxides of Fe (Coster and Kiestra), 144.
Fission, search for negative proton (Scott and Titterton), 918.
——, spontaneous from natural uranium (Whitehouse and Galbraith), 429.

Fluids, quantum mechanics (Price), 948.

Fluorescence of silver halides.—Pts. I. and II. (Farnell, Burton and Hallama), 157, 545.

Fourth order meson - equation (Thirring), 653.

Fowler (P. H.), *see* Camerini *et al.*, 413.

——, nuclear transmutations produced by cosmic ray particles of great energy, 169.

Frank (F. C.), lattice sums for ionic crystals, 1287.

——, structure of a crystal surface, 200.

Franzinetti (C.), mass of charged particles of cosmic radiation, 86.

Freeman (Miss J. M.), *see* Burcham and Freeman, 337, 873.

——, emission of short range alpha particles.—IV. $^{19}\mathrm{F}(p\alpha)^{16}\mathrm{O}$, 1225.

Frölich (H.), Pelzer (H.) and Zieman (S.), properties of slow electrons in polar materials, 221.

Fürth (R.), the "velocity" of brownian movement, 1255.

Gabor (D.), communication theory and physics, 1161.

Galbraith (W.), *see* Whitehouse and Galbraith, 429.

Gamma radiation, production of pairs (Hooper and King), 1194.

Gardner (F. F.), use of atmospherics to study propagation of very long radio waves, 1259.

Garrett (C. G. B.), thermal conductivity of potassium chrome alum below 1° K., 621.

Gas containing soundwaves, force on moving body (Nabarro), 1270.

—— discharge, energy distribution of electrons (Haigh and Smith), 557.

Geomagnetic radial variation, experimental determination (Runcorn, Benson, Moore and Griffiths), 783.

Germanium dioxide, decay of current (Tillman and Yemm), 1281.

Gething (P. J. D.), cut off in energy spectrum of primary cosmic rays, 495.

Gisolf (J. H.), back current in blocking-layer rectifiers, 754.

Goldsack (S. J.), *see* Dilworth, *et al.*, 1032.

Goldschmidt-Clermont (Y.), *see* Dilworth *et al.*, 1032.

——, fields of spherical symmetry (Narliker and Singh), 152.

Gravitational sphere of fluid, thermal equilibrium (Buchdahl), 362.

Green (A. E.) and Zerna (W.), theory of elasticity in general coordinates, 313.

Greenhow (J. S.), fluctuation and fading of radio echoes from meteor trails, 682.

Griffin (L. J.), observations on unimolecular growth steps on crystal surfaces, 196.

Griffiths (D. H.), *see* Runcorn, Benson, Moore and Griffiths, 783.

Gross (B.), inversion of Laplace transform, 543.

Guest (P. G.), orthogonal polynomials in least squares fitting of observations, 124.

Haigh (C.) and Smith (T.), energy distribution of electrons in positive columns, 557.

Hallama (R.), *see* Farrell, Burton and Hallama, 157, 545.

Harding (J. B.), further evidence for the existence of τ-mesons, 405.

Hardness, effect of specimen size in measurement (Hill), 745.

de Heer (J.), influence of overlap integral in molecular-orbital calculations for alternant hydrocarbons, 370.

Helium II, velocity of flow in creeping film (Jackson and Henshaw), 1081.

——, transfer rates for surface films (Burge and Jackson), 205.

Henshaw (D. G.), *see* Jackson and Henshaw.

Herdmann (J.), scattering of 90 MeV. neutrons by alpha particles, 444.

Hodson (A. L.), Loria (A.) and Ryder (N. V.), control of Wilson cloud chamber by means of an internal counter, 826.

Hodgson (P. E.), gap measurement of cosmic ray stars in emulsions, 725.

Hooper (J. E.), *see* Carlson *et al.*, 701.

Hooper (J. E.) and King (D. T.), multiple processes in the production of pairs of electrons by γ radiation, 1194.

Hill (R.), the inhomogeneous deformation of a plastic lamina in a compression test, 733.

——, theoretical investigati ns of the effect of specimen size in the measurement of hardness, 745.

——, theory of plastic bulging of a metal diaphragm by lateral pressure, 1133.

Insch (G. M.), *see* Curran *et al.*, 517.

——, calibration of proportional counters by excitation of fluorescence radiation, 857.

Interaction of collodial particles.—III. (Levine), 53.

Intrinsic block structure and the strength of metals (Paterson), 410.

Ionic crystals, lattice sums (Frank), 1287.

Ionized gas, refractive index and radiactive processes (Westfold), 509.

Ionizing particles, energy losses in photographic plate (Messell and Ritson), 1129.

Jackson (L. C.) and Henshaw (D. G.), the velocity of flow of He II in the creeping film, 1081.

—— and Preston-Thomas (H), superconductivity of Niobium, 1284.

—— *see* Burge and Jackson, 205.

Jacobs (J. A.), relaxation methods applied to plastic flow.—I. Notched bar under tension, 349.

—— .—II., 458.

Jelley (J. V.), the radioactivity of F^{20}.—I., 1199.

Jones (H.), effect of electron concentration on lattice spacing in magnesium solid solids, 663.

Kent (D. W.), *see* Dainton and Kent, 963.

Kiestra (S.), *see* Coster and Kiestra, 144.

King (D. T.), *see* Carlson *et al.*, 701.

——, *see* Hooper and King, 1194.

Laminar boundary layer on solids of revolution (Whitehead and Canetti), 983.

Laplace transform, inversion (Gross), 543.

Lattimore (S.), latitude effect of cosmic ray stars, 961.

Lattimore (S.), origin of single cosmic ray protons, 819.

Leah (A. S.), Rounthwaite (C.) and Bradley (D.), resistance thermometry in the study of gaseous explosions, 468, 1289.

—— ——, atomic oxygen in CO explosions, 478.

Levine (S.), interaction of colloidal particles.—III. Free energy of double layers, 53.

Levy (F.), *see* Dilworth *et al.*, 1032.

Li (T. T.), double stars due to cosmic rays, 1152.

Liebmann (G.), method of mapping of vector potential distributions in axially symmetrical systems, 1143.

Lighthill (M. J.), energy distribution behind decaying shocks, 1101.

Lilley (B. A.), energies and widths of domain boundaries in ferromagnetics, 792.

Littauer (R. M.), the radioactivity of F^{20}.—II., 1214.

Lock (W. O.), *see* Camerini *et al.*, 413.

Loria (A.), *see* Hodson, Loria and Ryder, 826.

MacDonald (D. K. C.), the statistical analysis of electrical noise, 814.

——, transit-time phenomena in electron streams.—II., 863.

MacKay (D. M.), quantal aspects of scientific information, 289.

Magnesium, cohesive energy of metallic (Raines), 568.

—— solid solutions, effect of electron concentrations (Jones), 663.

Magnetic deflection of fast charged particles in the photographic emulsion (Dilworth *et al.*), 1032.

—— spectrometer (Rae), 525.

Mapping vector potential distributions (Liebmann), 1143.

Matthews (P. T.), S-matrix for meson-nuclear interactions, 185.

Meltzer (B.), noise in current-carrying ohmic conductors, 393.

Menon (M. G. K.), Muirhead (H.) and Rochat (O.), reactions produced by slow negative π-mesons, 583.

Mesons, further evidence for the existence of τ-mesons (Harding), 405.

Mesons, reactions by slow negative π (Menon *et al.*), 583.
——, π nature of neutral particle in decay (O'Ceallaigh), 838.
Messel (H.) and Ritson (D. M.), energy losses of particles at relativistic velocities in the photographic plate, 1129.
Meteor trails, fluctuation and fading of radio echoes (Greenhow), 682.
——, measurement of height (Clegg and Davidson), 77.
—— trail ionization, influence of wind (Ellyett), 694.
Mixed ketones, order disorder transitions (Daniel), 631.
Moore (A. F.), *see* Runcorn, Benson, Moore and Griffiths, 783.
Moyal (J. E.), momentum and sign of fast cosmic ray particles, 1058.
Muirhead (H.), *see* Camerini *et al.*, 413.
——, *see* Menon *et al.*, 583.
Nabarro (F. R. N.), force on body moving uniformly through gas containing sound waves, 1270.
Narliker (V. V.) and Singh (K. P.), gravitational fields of spherical symmetry, 152.
Noise in ionization chamber pulse amplifiers (Wilson), 66.
——, in ohmic conductors (Meltzer), 393.
Nucleus, position of lowest dipole-level (Touschek), 849.
O'Ceallaigh, nature of neutral particle in decay of the π-meson, 838.
Orthogonal polynomials in lead squares (Guest), 124.
Osborne (D. V.), *see* Atkins and Osborne, 1078.
Papapetrou (A.), four dimensional generalization of Wilson's hypothesis, 399.
Paterson (M. S.), intrinsic block structure and strength of metals, 410.
Paul (E. B.), energy release in the reaction ^{7}Li(dp)^{8}Li, 942.
——, *see* Baxter, Burcham and Paul, 937.
Pelzer (H.), *see* Fröhlich, Pelzer and Zienan, 221.
Perkins (D. H.), emission of protons and alpha-particles in high energy cosmic ray stars, 138.
Pippard (A. B.), kinetics of phase transition in superconductors, 243.
Plastic flow, relaxation methods (Jacobs), 349, 458.
——, bulging of metal diaphragm (Hill), 1133.
Preston-Thomas (H.), *see* Jackson and Preston-Thomas, 1284.
Price (P. J.), quantum mechanics of fluids, 948.
Properties of slow electrons in polar materials (Fröhlich, Pelzer and Zienan), 221.
Proportional counters, calibration (Insch), 857.
—— ——, in magnetic field (West and Rothwell), 873.
Protons, fast, loss of energy in matter (Taylor), 966.
Radar information, theory (Woodward and Davies), 1001.
Radioactive corrections in non-relativistic limit (Thirring), 1193.
Radioactivity of F^{20}.—I. (Jelley), 1199.
—— ——.—II. (Littauer), 1214.
Rae (E. R.), magnetic spectrometer of large solid angle, 525.
Raimes (S.), cohesive energy of metallic magnetium, 568.
Rectifiers, back current in blocking-layer (Gisolf), 754.
Resistance thermometry in rapidly changing temperatures (Daish *et al.*), 729.
—— —— in gaseous explosions (Leah *et al.*), 468, 1289.
Richter (A. F.), the di-iso-electronic row and effective charge in L isorows, 983.
Ritson (D. M.), *see* Messel and Ritson, 1129.
Rochat (O.), *see* Menon *et al.*, 583.
Rolled sheet in tension and cupping tests, directional properties (Bourne and Hill), 671.
Rothwell (Miss P.), *see* West and Rothwell, 873.
Rounthwaite (C.), *see* Leah *et al.*, 468, 478, 1289.

Runcorn (S. K.), Benson (A. C.), Moore (A. F.) and Griffiths (D. H.), experimental determination of the geomagnetic radial variation, 783.
Ryder (N. V.), *see* Hodson, Loria and Ryder, 826.
Scattering of 90 MeV. neutrons by alpha particles (Heidmann), 444.
—— of pseudoscalar charged mesons (Corinaldesi and Field), 364.
Scientific information, quantal aspects of (Mackay), 289.
Scott (J. M. C.) and Titterton (E. W.), search for negative proton in fission, 918.
Second sound below 1° K., velocity (Atkins and Osborne), 1078.
Shannon (C. E.), programming a computor for playing chess, 256.
Singh (K. P.), *see* Narliker and Singh, 152.
S-matrix for meson-nuclear interactions (Matthews), 185.
Smith (T.), *see* Haigh and Smith, 557.
Snoek (J. L.), density variations in aluminium, 1188.
Soft radiations VII. (Curran, Cockroft and Insch), 517.
Soh (H. P.), calculation of energy states by uncertainty relations, 851.
Sound waves, propagation in an open-ended channel (Chester), 11.
Specific heat of free electrons in metals (Wohlfarth), 534.
Statistics of correlated events.—I. (Domb), 969.
Structure of a crystal surface (Frank), 200.
Superconductors, kinetics of phase transition (Pippard), 243.
Superconductivity of Niobium (Jackson and Preston-Thomas), 1284.
Taylor (A. E.), loss of energy of fast protons in matter, 966.
Tempest (R. K.), velocities and local Mach numbers in flow of compressible fluid, 382.
Thermal conductivity of potassium chrome alum below 1° K. (Garrett), 621.
Thirring (W.), fourth order meson-equation, 653.
——, radiative corrections in non-relativistic limit, 1193.
Thorum-232, binary and ternary photofission (Titterton and Brinkley), 500.
Tillman (J. R.) and Yemm (H.), decay of current in germanium dioxide, 1281.
Titterton (E. W.), *see* Scott and Titterton, 918.
—— and Brinckley (T. A.), binary and ternary photofission of Th-232, 500.
Touschek (B. F.), position of lowest dipole-level of a nucleus, 849.
Townsend (A. A.), eddy viscosity in turbulent shear flow, 890.
Unimolecular growth steps on crystal surfaces (Griffin), 196.
"Velocity" of brownian movement (Fürth), 1255.
West (D.) and Rothwell (Miss P.), proportional counters in magnetic field; isomerism of Br^{80}, etc., 873.
Westfold (K. C.), refractive index and classical radiative processes in an ionized gas, 509.
Whitehead (L. G.) and Canetti (G. S.), the laminar boundary layer on solids of revolution, 988.
Whitehouse (W. J.) and Galbraith (W.), energy spectrum from spontaneous fission of natural uranium, 429.
Wilkins (J. J.), betatron-starting in electron sychrotrons, 34.
Wilson cloud chamber, control (Hodson, Loria and Ryder), 826.
Wilson (R.), noise in ionization chamber pulse amplifiers, 66.
Wilson's hypothesis, four dimensional generalization (Papapetrou), 399.
Wohlfarth (E. P.), influence of exchange and correlation forces on specific heat of electrons in metals, 534.
Woodall (A. J.), *see* Daish *et al.*, 729.
Woodward (P. M.) and Davies (I. L.), theory of information, 1001.
Yemm (H.), *see* Tillman and Yemm, 1281.
Zerna (W.), *see* Green and Zerna, 313.
Zienan (S.), *see* Fröhlich, Pelzer and Zienan, 221.

END OF THE FORTY-FIRST VOLUME.

Printed in the Netherlands